DISSECTION OF THE MECHANISMS UNDERLYING GENOME-WIDE TRANSCRIPTIONAL REGULATION OF MAMMALIAN STRESS RESPONSE

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

In Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Dig Bijay Mahat

January 2017

DISSECTION OF THE MECHANISMS UNDERLYING GENOME-WIDE TRANSCRIPTIONAL REGULATION OF MAMMALIAN STRESS RESPONSE

Dig Bijay Mahat, Ph. D.

Cornell University 2017

Variation in growth conditions is an inevitable consequence of the environment's dynamism and, as such, organisms are subjected to adverse living conditions including non-optimal temperature. Exposure to elevated temperature and other proteotoxic stresses disrupts structural integrity of cellular machineries and results in cytotoxic unfolded-protein-aggregates. To cope with proteotoxic stresses, organisms are equipped with an evolutionarily conserved defense mechanism known as heat shock response (HSR). Despite the significance of HSR in protein homeostasis and survival, it's scope, extent, and the molecular mechanism of regulation are poorly understood.

Our work shows that the genome-wide transcriptional response to heat-stress in mammals is rapid, dynamic, and results in induction of several hundred and repression of several thousand genes. Heat shock factor 1 (HSF1), 'the master regulator' of the HSR, controls only a fraction of the heat-stress induced genes, and does so by increasing RNA polymerase II (Pol II)

release from its promoter-proximal pause. The pervasive repression of transcription is predominantly HSF1-independent, and is mediated through reduction of Pol II pause-release. The up- and down-regulated genes during HSR are accompanied by concomitant increase and decrease respectively in promoter occupancy of pause-release factor positive transcription elongation factor b (P-TEFb). HSF2, the ubiquitously expressed paralog of HSF1, has a broader repressive role during stress, likely mediated through other factors, and its promoter-binding activity is dependent on HSF1. Our work also demonstrates the unprecedented role of serum response factor (SRF) in transient induction of cytoskeletal genes during the early phase of HSR. Overall, mammalian cells orchestrate rapid, dynamic, and extensive changes in transcription upon heat-stress that are largely modulated at pause-release, and HSF1 plays a limited and specialized role.

BIOGRAPHICAL SKETCH

Dig Bijay Mahat was born and raised in a rural village in the foothills of the Himalayas, 100 miles west of Kathmandu, Nepal. He grew up with three siblings and nine cousins in a large household. Having not completed the middle school, his parents were obessed with providing Dig a good education. They transferred him to nine different institutions by the end of high school. He was six years old when he walked 90 minutes everyday to attend first grade.

As a child, he enjoyed playing sports but did not excel in any of them. Instead he turned to science. Unable to complete the 11^{th} grade due to a health issue, he wrote poems and songs during recovery and shifted his focus from study to music. However, the failure of his music career reverted him to science once again. Motivated by scientific research, he defied his family's wish to go to a medical school in Nepal and instead came to the United States.

After a year in Salem University, he transferred to Towson University to study biology. It was in Dr. Gerald Wilson's lab in University of Maryland where he discovered his passion for biomedical research. The diagnosis of his mother with breast cancer in 2007 further motivated him to pursue cancer research. In 2010, he came to Cornell for graduate studies and joined Dr. John Lis' lab. Since the very first day, he was captivated by John's big vision and was drawn towards high-risk high-reward projects. John's influence has been the most profound experience of his life and he continues to learn from him.

बुवा, गुन्टिराम, र लिखोप्रति समर्पित

ACKNOWLEDGMENTS

I think of my academic career as a culmination of concerted efforts of my family, colleagues, and mentors. My father's determination to provide me with good education and my mother's grace and grit in cultivating a winning attitude is the bedrock of my academic foundation. Support from my extended family in the form of motivation and means has been momentous, and the expectation from my brothers and sisters either by raising the bar or by idealizing has provided fuel to excel.

Having gone through twelve different academic institutions, I have met many wonderful folks along the way who have become life-long friends. Through laughter and tears, and discussion and beers, they have left an indelible mark in my heart and I cherish each and everyone's friendship. But the one to live through the joys and challenges of my scientific career is my wife who is also my closest friend, harshest critic, and loudest cheerleader.

I have been extremely fortunate to have incredible mentors throughout my academic career. Some taught me to read and write, others taught me to think and fight, and they came at different phases of my life. But the one who I met at the end left the most lasting influence. I have learned science, philosophy, grammar, intramedic tubing, and the art of living from him.

I am extremely grateful to all of you. I am because of you.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

MECHANISM AND MACHINERY OF TRANSCRIPTIONAL REGULATION OF HEAT SHOCK RESPONSE

Summary

The dynamic nature of environment has a profound effect on evolution of life. Variation in the growth environment is an inevitable consequence of environment's dynamism, and as such, organisms are subjected to adverse living conditions including non-optimal temperature. Exposure to high-temperature disrupts the structural integrity of cellular proteins leading to formation of cytotoxic unfolded protein aggregates. To cope with elevated temperature, organisms have developed a cellular response mechanism known as heat shock response (HSR). This adaptive defense mechanism is evolutionarily conserved across the domains of life. The focus of this chapter is to provide an evolutionary context for the emergence of HSR, elaborate the roles of heat shock proteins in restoration of protein homeostasis, and the mechanisms of heat shock transcription factor-1 activation and its role in transcription regulation. Understanding the mechanism and machinery of HSR is critical in biology and diseases. The relentless pursuit of how major faculties of gene regulation contribute in its orchestration from studies in various species has unveiled the enormity and complexity of the HSR.

Introduction

Heat shock response (HSR) was first observed more than five decades ago as chromosomal puffs in *Drosophila* salivary glands (Ritossa, 1962). The discovery was considered a cellular artifact and struggled for some time to be embraced as a manifestation of a gene regulatory mechanism (De Maio et al., 2012). However, the detection of heat shock proteins (HSPs) in 1974 (Tissiéres et al., 1974) and the subsequent volumes of work in 80's and 90's on HSPs established the HSR as a critical and universal cellular defense mechanism. The stress response field has exploded ever since and the discovery of HSR in virtually every organism has catapulted it to be accepted as a general feature of cells. Similarity in the HSR among various organisms spans from instigating cues to activating molecules, and from mechanistic steps to synthesis of final products. Although HSR was discovered as a protective mechanism against heat stress, it is evoked by various stresses including toxins (Aït-Aïssa et al., 2000; Hensler et al., 1991; Liu et al., 1996; Löw-Friedrich and Schoeppe, 1991; Wirth et al., 2002), heavy metals (Boone and Vijayan, 2002; Deane and Woo, 2006; Singer et al., 2005; Warchałowska-Sliwa et al., 2005), oxidants (Moraitis and Curran, 2004; Trott et al., 2008; Wallen et al., 1997), radiations (Lin et al., 1997; Trautinger et al., 1996), metabolites (Abe et al., 1998; Chang et al., 2001), and pathogens (Kantengwa and Polla, 1993; Ramaglia et al., 2004). Furthermore, HSR can also be triggered by exercise (Fehrenbach and Niess, 1999; Leon, 2016; Yamada et

al., 2008), and nutrients (Mahat and Lis, 2016). As such, HSR is now regarded as the integrated cellular response for its ability to sense and respond to various stresses.

The defining features of HSR are well established. Despite the diversity in origin of stresses, they generate non-native proteins in cells and disrupt protein homeostasis (proteostasis) (Morimoto, 1998a). Disruption in proteostasis affects several aspects of cellular structure and function (Figure 1.1) and subsequently results in activation of the heat shock transcription factor 1 (HSF1) (Parker and Topol, 1984). Vertebrates have four *HSF* genes, *HSF1–4*. HSF1 is considered the 'master regulator' of HSR and is the ortholog of the sole *HSF* gene in invertebrates. Upon stress, the dormant monomeric-HSF1 trimerizes, binds to heat shock element (HSE, inverted repeats of nGAAn pentamers) in the promoters of heat shock protein genes (*HSPs*), and increases their transcription considerably (Perisic et al., 1989; Westwood et al., 1991; Xiao and Lis, 1988). Subsequent post-transcriptional and translational regulatory steps further accentuate *HSPs'* expression, whereas non-critical genes' expression is suppressed by coordination of the transcriptional, post-transcriptional, and translational regulatory steps.

Despite the flood of information on several aspects of the HSR, including various nature of activating cues and the stress-specific synthesis of *HSPs* (Roccheri et al., 2004), many features of this response mechanism remain poorly understood. How cells sense heat and other kinds of stresses?

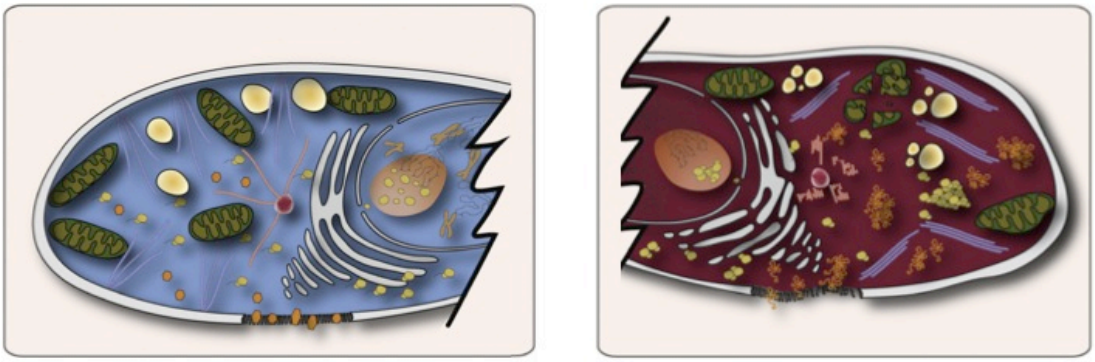


Figure 1.1. Deformation of cellular structures induced by heat stress
Heat stress damages cytoskeleton, including the reorganization of actin filaments (blue) into stress fibers and the aggregation of other filaments. Organelles such as Golgi and endoplasmic reticulum (white) become fragmented the number and integrity of mitochondria (green) and lysosomes (yellow-white gradient) decreases. Large depositions of the stress granula (yellow), resulting from assemblies of proteins and RNA, are found in the cytosol. Changes in the membrane morphology, aggregation of membrane proteins, and an increase in membrane fluidity also occur. Figure and description adapter from (Richter et al., 2010).

What triggers the activation of HSF1? What are the genes regulated by HSF1and what factors and mechanisms induce non-*HSPs*? How is global repression of non-essential genes achieved? After decades of research, some models and hypotheses have emerged that explain key observations in the HSR. However, advent and application of new tools routinely challenge the established models and new technologies continue to broaden our understanding of the causes and consequences of proteotoxic stress. The enduring exploration of mechanism and machinery of HSR is more important than ever because of its direct involvement in various diseases. This chapter focuses on the evolutionary context and functions of HSPs, the regulation of HSR with an emphasis on transcriptional regulation and the role of HSF1, and the implications of HSR in health and diseases.

Evolution of protein folding machinery

The foundation for the evolution of protein folding machinery is multifaceted. Many small cellular proteins have the ability to refold into their native three-dimensional structure with sub-second kinetics (Kubelka et al., 2004), indicating the structural information of a protein is encoded in its DNA (Dobson et al., 1998). However, folding of larger proteins may take from minutes to even hours (Herbst et al., 1997).The longer folding kinetics and a high *in vivo* protein concentration (300-400 ng/ul) (Dobson et al., 1998) makes larger proteins vulnerable for misfolding and thus requires assisting machinery

for proper folding. Furthermore, chaperone machinery serves as a capacitor of change in protein structure that arise from nonsynomous substitution of amino acids and mediates protein evolution by temporarily masking the phenotype of altered protein structure (Rutherford and Lindquist, 1998). Lastly, extracellular and intracellular cytotoxic agents can potentially alter the free-energy landscape of cellular proteins, which makes proteostasis restoration mechanism a critical requisite for continued cellular functions (Morimoto, 1998a). These scenarios that were likely encountered at the dawn of evolution are responsible for the ubiquity of protein folding machinery.

Heat shock proteins in proteostasis

The objective of HSR from sensing ectopic stress to synthesizing HSPs by employing multiple layers of regulatory events from transcription to translation is to restore proteostasis. If HSR is an emergency response, HSPs are the paramedics, and there seems to be a variety of paramedics with unique specializations. Some HSPs serve as the hub of protein refolding machinery whereas others recognize and deliver the “injured” proteins to the refolding machinery. In general, HSPs utilize ATP molecules to restore the native conformation of unfolded and folding-intermediate proteins with the help of co-chaperones. Moreover, different HSPs characterize different diseases invoking the concept of HSP barcoded diseases (Kakkar et al., 2014).

HSP110 family members are close relatives of HSP70 with a longer linker region between the N-teminal ATPase domain and the C-terminal peptide-binding domain resulting in their larger size (Easton et al., 2000). These proteins associate with and assist the protein folding activity of HSP70 by serving as a nucleotide exchange factor during ATP hydrolysis in chaperoning cycle (Andréasson et al., 2008; Dragovic et al., 2006; Raviol et al., 2006). This auxiliary role of HSP110s in HSP70-mediated protein folding and their high efficiency and differential preference in substrate holding have led some scientists to argue that they lack chaperoning activity (Zuo et al., 2016), however, studies using heat-denatured luciferase (Yamagishi et al., 2011) and misfolded and aggregated polypeptides (Mattoo et al., 2013) as substrates have provided evidence for the direct role of HSP110s in protein folding and established them as bona fide chaperones. In yeast, HSP110 is also critical for acquired thermotolerance (Sanchez and Lindquist, 1990).

HSP110s have implications beyond the protein-folding space. They mitigate oxidative stress-induced apoptosis through p38 MAPK-dependent mitochondrial pathway (Yamagishi et al., 2008). Targeting of HSP110 by small synthetic interfering RNA induces apoptosis of various cancer cells indicating addiction of cancer cells to HSP110s (Hosaka et al., 2006). Interestingly, they also play a critical role in innate immunity by delivering antigens to the endogenous antigen-presenting pathway for cross-presentation and thus are

being considered for cancer immunotherapy (Bolhassani and Rafati, 2008; Murshid et al., 2008).

Hsp90s are another major class of HSPs. They are highly abundant dimeric protein constituting 1-2% of total cellular protein content. Eukaryotic genomes encode multiple compartment-specific HSP90 proteins and their conservation across species indicate that they likely arose early in evolution (Taipale et al., 2010). Despite the high intrinsic flexibility of HSP90, its structural properties in bacteria, yeast, and mammals have been resolved, which indicate that the HSP90s share structural homology with other ATPases (Ali et al., 2006; Dollins et al., 2007; Shiau et al., 2006). The highly conserved amino-terminal domain binds ATP and the carboxy-terminal domain mediates dimerization (Csermely et al., 1993). ATP binding triggers a conformational change in HSP90 causing a transition from an open (ADP-bound) to a closed (ATP-bound) form (Grenert et al., 1997), and induces structural readjustments necessary for client protein's conformational change (Chadli et al., 2000). However, recognition of client proteins by HSP90 is still a mystery. Its diverse clientele lack structural or sequence similarity, prompting the model where client proteins are delivered to HSP90 by other chaperones and co-chaperones. Indeed, HSP70 interacts with HSP40 to recognize short hydrophobic exposed segment of nascent polypeptides or partially unfolded proteins, which are brought to HSP90 for refolding (Genevaux et al., 2007; Rüdiger et al., 1997).

Besides assisting in the folding of nascent polypeptide and partially-unfolded proteins, HSP90, in association with other chaperones and co-chaperones, sequesters HSF1 in a repressive complex (Zou et al., 1998). HSP90 also cooperates with HSP70 to transport a subset of mitochondrial preproteins for subsequent membrane translocation (Young et al., 2003). It capacitates alteration in protein structure and facilitates evolution by temporarily masking cryptic variation (Jarosz and Lindquist, 2010). Because of its role in proteostasis, cancer cells heavily rely on HSP90 to stabilize mutant proteins that would otherwise be degraded (Li et al., 2011; Trepel et al., 2010). Predictably, several HSP90 inhibitors designed for cancer therapeutics are at different phases of clinical trial (Neckers and Workman, 2012).

Regulation of heat shock response

The regulation of HSR is an intricate process accomplished by engaging various faculties of regulation. Cells deploy transcriptional (Duarte et al., 2016; Mahat et al., 2016; Solís et al., 2016), post-transcriptional (Meyer et al., 2015; Shalgi et al., 2014; Zhou et al., 2015) , translational (Shalgi et al., 2013) , and post-translational (Raychaudhuri et al., 2014) programs to regulate the HSR. Feedback regulation between translational and transcriptional processes further complicates the regulatory cycle (Santagata et al., 2013). The evolution of checks and balances ensures optimal expression of chaperone proteins and temporal control of their expression, whereas, the

tight coordination between various regulatory steps underlies the significance of this cellular response mechanism in survival.

So why do cells conjure such concerted effort from all major regulatory branches of gene expression in response to stress? The answer is rather simple – stress is harmful, even lethal in case of increased severity or duration, and cells do everything to promote their survival. In order to survive, cells halt non-critical functions, conserve energy, and selectively mobilize pro-survival programs. The rapid and robust expression of pro-survival genes and suppression of non-critical functions is a keenly regulated process - it begins with instantaneous sensing of heat stress by cells and continues all the way to recovery from heat stress.

Mechanisms to sense heat stress and activation of HSF1

The field of stress response has taken significant strides in identifying the events downstream of HSF1 activation and the subsequent changes in transcriptional and translational programs caused by proteotoxic stress. However, after more than five decades since the discovery of HSR, the primary mechanism and factors responsible to sense heat stress remain elusive. The quest to identify the pioneering events in incitation of the HSR is further complicated by the discovery of several distinct models of heat sensing and HSF1 activation. The three predominant models in the field of HSR are discussed below.

Chaperone titration

Release of inert monomeric HSF1 from chaperone-associated complex is the predominant model of HSF1 activation. The influx of protein folding intermediates due to heat stress is sensed by HSP90, HSP70, and their co-chaperones and the chaperones dissociate from the HSF1-interacting complex liberating HSF1 to trimerize and activate (Morimoto, 1998b; Neef et al., 2014). This model is supported by the activation of HSF1 and subsequent expression of *HSPs* triggered by the injection of denatured proteins into *Xenopus* oocytes but not by native proteins (Ananthan et al., 1986). Furthermore, multiple observations of interaction between the chaperones and HSF1 (Abravaya et al., 1992; Duina et al., 1998; Nadeau et al., 1993; Neef et al., 2014; Ohama et al., 2016) indicate the presence of such repressive complex. This interaction is reduced during proteotoxic stress (Guo et al., 2001), and pharmacological inhibition of HSP90 releases HSF1 to trimerize and bind to HSE *in vitro* (Zou et al., 1998). A more recent study using experimental and theoretical approaches tested this model and found that the dynamic interaction of HSP70 with HSF1 forms the basis of HSF1 activation switch (Zheng et al., 2016). The influx of unfolded proteins caused by stress titrates away HSP70 enabling HSF1 to trimerize and activate. To this date, this is the most compelling evidence for the activation of HSF1 via titration of chaperones from the inactive complex. Nevertheless, the existing models of HSF1 activation are not necessarily mutually exclusive and other models could still contribute in heat sensing.

Post-translational modification of HSF1

Post-translational modifications of HSF1 serve as another putative mechanism for cell's ability to sense heat stress and activate HSF1. This model emerged from two sets of observations. First, the release of latent monomeric HSF1 from repressive chaperone complex and subsequent trimerization and binding to HSEs is not sufficient to induce transcription, suggesting the requirement of further activating cues in HSF1-mediated transcription induction (Cotto et al., 1996; Jurivich et al., 1995; Petesch and Lis, 2008). Second, HSF1 is heavily phosphorylated during heat stress suggesting that the post-translational modifications may serve as the ultimate activating cue that renders HSE-bound HSF1active (Guettouche et al., 2005; Holmberg et al., 2001; Lee et al., 2008; Soncin et al., 2003; Sorger and Pelham, 1988). However, phosphorylation of some serine residues are associated with HSF1 repression (Høj and Jakobsen, 1994; Kline and Morimoto, 1997; Knauf et al., 1996) and phosphorylation-dependent sumoylation also has an inhibitory role on HSF1 function (Hietakangas et al., 2003). Furthermore, acetylation of the lysine residues of HSF1 by p300 serves as a negative regulator of HSF1's DNA binding ability, whereas SIRT1-mediated deacetylation maintains HSF1 in a DNA-binding competent state (Westerheide et al., 2009).

The surprising finding regarding the post-translational modifications of HSF1 emerged when the mutagenesis of all known phosphoacceptor residues

of HSF1 did not hamper its transcription activation potential (Guettouche et al., 2005). Simultaneous mutagenesis of 15 phosphorylation sites had no effect on HSF1's activation (Budzyński et al., 2015) and genome-wide RNAi screens failed to identify kinases as modulators of HSF1 activity (Raychaudhuri et al., 2014). Nevertheless, quantitative examination of the contribution of phosphorylation revealed its dispensability on HSF1 activation and revealed its role as a 'fine tuning' knob on HSF1 activity (Zheng et al., 2016).

Intrinsic ability of HSF1 to sense stress

HSF1's intrinsic ability to sense heat adds another dimension to the heat sensing landscape during stress. Monomeric HSF1 can trimerize *in vitro* by heat stress, oxidative stress, low pH, and higher calcium concentration (Goodson and Sarge, 1995; Mosser et al., 1990; Zhong et al., 1998). HSF1's thermosensory ability is manifested by temperature-dependent unfolding of the heptad region in its regulatory domain (HR-C) and concomitant stabilization of oligomerization domain (HR-A/B) (Hentze et al., 2016). Substitution of three hydrophobic residues in the HR-C domain renders HSF1 constitutively trimerized and competent for DNA binding (Rabindran et al., 1993; Zuo et al., 1994). The trimerized state of human HSF1 achieved through this genetic manipulation is capable of translocating into the nucleus when expressed in yeast cells, while native human HSF1 localizes throughout the yeast cell, and complements the viability defect associated with the deletion of yeast HSF

gene (Liu et al., 1997). Besides the aforementioned evidence of temperature-dependent unfolding of HSF1, the inherent ability of HSF1 in heat-sensing is poorly understood. Nevertheless, the built-in sensory capability of HSF1 provides a plausible explanation for the promptness of the HSR activation, which is difficult to reconcile with the first two models that require separate molecular events to occur prior to HSF1 activation.

Beyond the three leading models of HSF1 activation, anecdotal evidences indicate the existence of two additional mechanisms of HSF1 activation. First, a ribonuclear protein complex that consist heat shock RNA-1 (HSR-1) facilitates HSF1 activation (Shamovsky et al., 2006). This model suggests that the HSR-1 serves as a thermosensor by undergoing heat-induced conformational change (Kugel and Goodrich, 2006). Second, a heat-stress dedicated neuronal circuit controls the HSR in *C. elegans* (Prahlad et al., 2008). Such neuronal network is yet to be identified in mammals, nevertheless, the finding in *C. elegans* indicates a likely organismal level control in stress response. Overall, organisms deploy several mechanisms that are not necessarily mutually exclusive to transduce stress signal to HSF1, which in turn activates highly coordinated transcriptional programs.

Transcriptional regulation of heat shock response

Once HSF1 is activated by heat stress, it orchestrates a transcriptional program designed to maximize cell's survival, the molecular details of which

are just beginning to emerge. Historically, study of the HSR was limited to the focused gene studies of *HSPs*. The confined window of HSF1 and HSPs and the lack of genome-wide tools masked the extent and diversity of the transcriptional program regulated during HSR for a long time. As a result, HSF1 has been considered the 'master regulator' of HSR. However, development of highly sensitive genome-wide methods has enabled the discovery of the diversity of transcriptional programs regulated during HSR.

HSF1 in HSR

It is well-known that HSF1 transcriptionally induces *HSPs* in all model organisms (Clos et al., 1990; Lindquist, 1986; McMillan et al., 1998). Much of what we know today about the basic mechanism of transcription regulation of *HSPs* by HSF1 comes from the focused gene studies in yeast (Gallo et al., 1993; Lindquist, 1986; Lindquist and Kim, 1996; Sorger and Pelham, 1988; Verghese et al., 2012; Wiederrecht et al., 1988) and *Drosphila Hsp70* gene (Giardina and Lis, 1995a; Jedlicka et al., 1997; Lis et al., 1983; Perisic et al., 1989; Rougvie and Lis, 1988; Zobeck et al., 2010).

Essentiality of HSF1

Acute inactivation of HSF1 through the "anchor-away" approach even during normal growth condition triggers toxic protein aggregates in yeast (Solís et al., 2016). This observation indicates the requirement of HSF1 in

maintaining protein homeostasis and corroborates the essentiality of yeast HSF1 for cell growth and viability (Gallo et al., 1993; Sorger and Pelham, 1988; Wiederrecht et al., 1988). Mutations disrupting DNA binding domain or activation domain of yeast HSF1 is lethal even in the unstressed condition (Jakobsen and Pelham, 1991; Torres and Bonner, 1995). In contrast, HSF1 is not essential for normal growth and viability in *Drosophila* but is required during larval development (Jedlicka et al., 1997). Similarly, mammalian cells can survive without HSF1 as long as they don't encounter stress and their lifespan is uncompromised (McMillan et al., 1998). However, lack of HSF1 can severely reduce their survival likelihood in event of stress (Luft et al., 2009). The reliance of mammalian cells on HSF1 only during stress indicates an alternative mechanism exists for basal expression of *HSPs*. It further suggests the adaptation of HSF1 as a stress-specialist transcription factor in mammalian cells.

Structure of HSF1

Structure of a biological molecule can convey vital information on its functional characteristics. HSF1 protein ranges from ~50-90 kDa in different species and is composed of structurally distinct domains. The winged helix-turn-helix DNA binding domain (DBD) lies in the N-terminal, followed by coiled-coil trimerization domain - a largely unstructured regulatory domain that consists a short heptad repeat, and a transactivation domain that is also

generally unstructured (Wu, 1995). Molecular structure of the DBD (Harrison et al., 1994) and DBD-HSE complex (Littlefield and Nelson, 1999) of yeast, DBD of *Drosophila* (Vuister et al., 1994), DBD (Liu et al., 2011) and DBD-HSE (Neudegger et al., 2016) of human HSF1 have been resolved so far. The most recent and comprehensive analysis of HSF1 structure yet showed the DBD of HSF1 completely embracing the DNA (Neudegger et al., 2016), providing an ample opportunity for other factors to interact with DNA-bound HSF1. It also revealed the conspicuous lack of deep pockets for small molecule drug to bind and interfere HSF1's DNA binding ability (Neudegger et al., 2016).

DNA binding ability of HSF1

Another critical feature in HSF1-mediated transcription regulation is its ability to bind DNA. Long stretches of DNA are condensed into chromatin using nucleosomes. Some transcription factors, known as 'pioneer factors', are able to bind condensed chromatin, recruit co-factors and chromatin remodelers to open local chromatin to create conducive environment for transcription initiation (Zaret and Carroll, 2011). Non-pioneer transcription factors bind to already accessible chromatin sites only. Some studies have suggested a likely pioneering transcriptional activity of HSF1. HSF1 opens the chromatin structure of Interleukin-6 promoter (Inouye et al., 2007) and stimulates demethylation of the promoter to facilitate p65 and c-Jun binding (Rokavec et al., 2012). Association with replication protein A (RPA) and

subsequent recruitment of histone Chaperone FACT allows HSF1 to access nucleosomal DNA (Fujimoto et al., 2012). However, conclusive experiment showing pioneering activity of HSF1 in more than few genomic loci remains to be conducted. Meanwhile, observations of active chromatin marks such as histone acetylation and H3K4 trimethylation dictating the heat-inducible binding of HSF1 to DNA (Guertin and Lis, 2010) underminess the hypothesis of HSF1 being a pioneer factor. Alteration of chromatin landscape from inactive to active state permits inducible HSF1 binding (Guertin and Lis, 2010), further indicating HSF1's inability to overcome inactive chromatin landscape.

While HSF1's ability to access chromatin remains to be fully known, its DNA binding mechanics is fairly understood. HSF1 binds DNA mostly as a trimer with multiple dyad symmetries and the trimerization is stabilized through coil-coin interaction (Perisic et al., 1989). DNA bound HSF1 trimer is further stabilized by intermolecular disulfide bond between individual monomers (Ahn and Thiele, 2003). Interference of the disulfide bond by reducing agents minimizes trimerization and stress-responsive gene transcription (Neef et al., 2010a). Once bound, it stably associates with DNA and facilitates multiple cycles of transcription (Yao et al., 2006).

Interacting partners and recruited factors

HSF1 binding to promoters is not sufficient to induce transcription suggesting the requirement of additional events (Cotto et al., 1996; Giardina

and Lis, 1995b; Jurivich et al., 1995; 1992; Petesch and Lis, 2008). One possible additional event is the modification of DNA-bound HSF1 by interacting partners or recruited factors that confers competency for transcription activation. Upon stress, P-TEFb kinase is recruited to heat shock loci in HSF1-dependent manner as shown by polytene spreads (Lis et al., 2000) and live cell imaging (Zobeck et al., 2010). Small molecule mediated inhibition of HSF1's ability to recruit P-TEFb results in lack of *HSPs* induction (Yoon et al., 2011). Similarly, a direct interaction between HSF1 and a mediator subunit recruits co-activating mediator complex to heat shock loci (Kim and Gross, 2013; Park et al., 2001). Genome-wide screens of HSF1 interacting partners through co-immunoprecipitation assay have identified more than 30 proteins that play a role in DNA repair, splicing, signaling, and protein degradation among others (Fujimoto et al., 2012; Kang et al., 2015). Given the promiscuous nature of co-immunoprecipitation assay, whether these interactions are functional remains to be shown.

HSF1 also modifies chromatin landscape to facilitate transcription. Stimulation of histone acetyl transferase Tip60 by HSF1 and consequent histone H2A lysine 5 acetylation boosts transcription activation of Drosophila *Hsp70* gene (Petesch and Lis, 2012). HSF1 recruits histone actetyltransferase p300/CBP to HSF1 bound sites (Hong et al., 2004; Smith et al., 2004), which in-turn acetylates HSF1 (Raychaudhuri et al., 2014). However, HSF1 acetylation is a dynamic process and the modulation of this mark has a

potential to alter HSF1's function. As discussed earlier, SIRT1 mediated deacetylation of HSF1 is required for its ability to bind DNA and inhibition of SIRT1 catalytic activity attenuates HSF1-DNA interaction and the HSR (Westerheide et al., 2009).

Mechanism of HSF1-mediated transcription induction

Transcription cycle consists of multiple rate-limiting steps (Fuda et al., 2009). Transcription factors often control either one or both of the two early steps in transcription cycle - Pol II recruitment and release of Pol II from the promoter-proximal pause (Adelman and Lis, 2012). The contribution of HSF1 in recruitment and/or release of Pol II was unclear until the development of genome-wide nuclear run-on assays. Precision run-on and sequencing (PRO-seq) is a highly sensitive genome-wide assay that maps RNA polymerase at base-pair resolution and thus provides quantitative transcriptional status. Use of PRO-seq in HSF1 knock out mouse embryonic fibroblasts revealed that the HSF1 induces transcription by increasing the release of Pol II from promoter-proximal pause (Duarte et al., 2016; Mahat et al., 2016). The recruitment of Pol II also increases during HS resulting in net increase in paused Pol II; however, the increase in Pol II recruitment and pausing is independent of HSF1 (Mahat et al., 2016). The primary candidate factor in mediating HSF1-induced pause release is P-TEFb kinase. Earlier observations of HSF1-dependent P-TEFb recruitment to heat shock loci indicate cooperation

between the two factors (Lis et al., 2000; Zobeck et al., 2010). Quantitative assessment of the occupancy of Cyclin dependent kinase 9 (Cdk9), the kinase subunit of P-TEFb, show that the P-TEFb is significantly enriched in the promoters and proximal gene body regions of HSF1-dependent genes during HSR (unpublished data, see Chapter 2). It further supports the existing model of HSF1 inducing transcription by accelerating the release of Pol II from promoter-proximal paused state (Figure 1.2).

Another defining feature of the HSR is the global downregulation of genes. More than 50% of actively transcribing genes are significantly downregulated during HSR (Mahat et al., 2016). The global repression of transcription is attained by the inhibition of Pol II release form promoter-proximal pause (Mahat et al., 2016). The reduction in P-TEFb occupancy in the promoters of repressed genes (unpublished data, see Chapter 2) provides a likely mechanism where reduction in P-TEFb availability impedes the maturation of paused Pol II into productive elongation complex (Figure 1.3). Together, the dependence of upregulated genes on P-TEFb and the deficit of P-TEFb in downregulated genes support a comprehensive model of gene regulation during HSR. At HSF1 induced genes, HSF1 recruits P-TEFb to increase the rate of Pol II release from the promoter-proximal pause. The shuttling of P-TEFb to HS-induced genes by HSF1 and possibly other transcription factors titrates P-TEFb away from genes that will eventually be repressed. Fluorescent recovery after photobleaching (FRAP) analysis

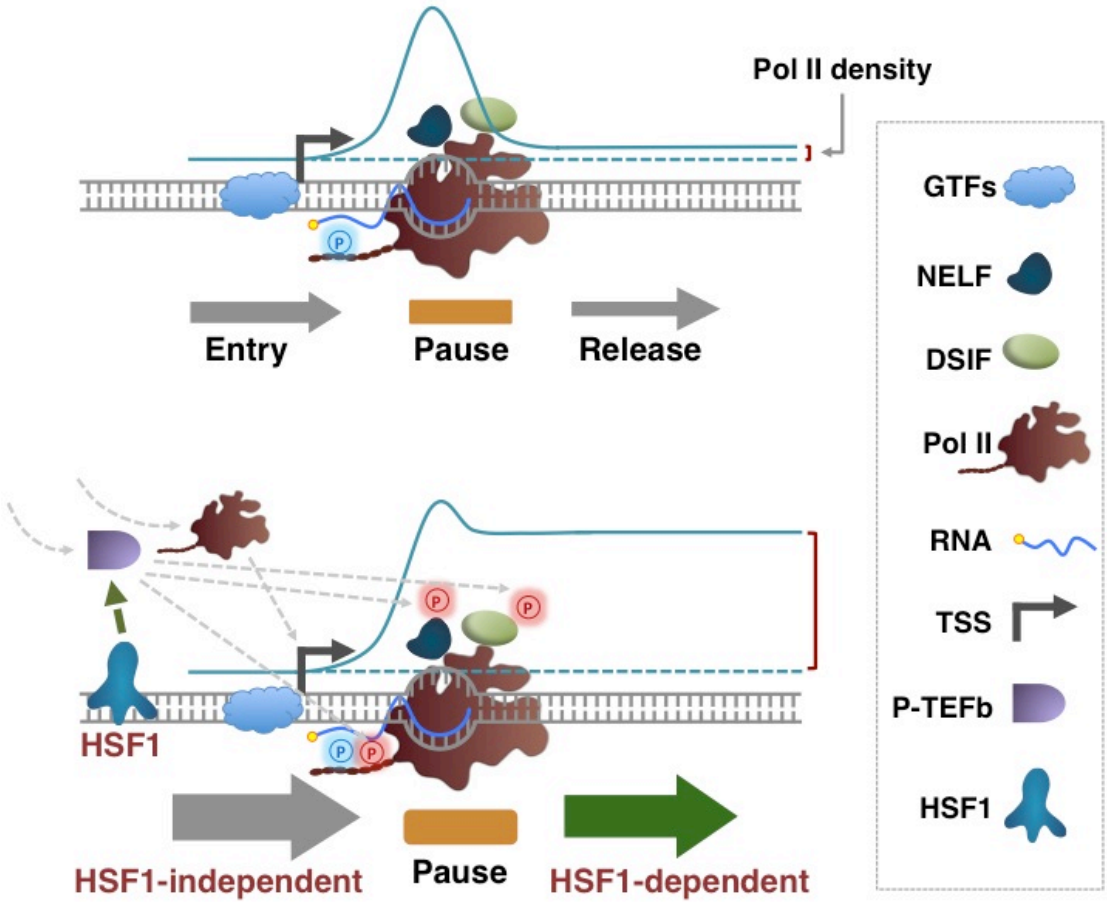


Figure 1.2. Mechanism of HSF1-mediated transcription induction during HSR

HSF1 binds and recruits the pause release factor P-TEFb, which accelerates the release of paused Pol II from the promoter-proximal paused state. The recruitment of Pol II, however, occurs independent of HSF1.

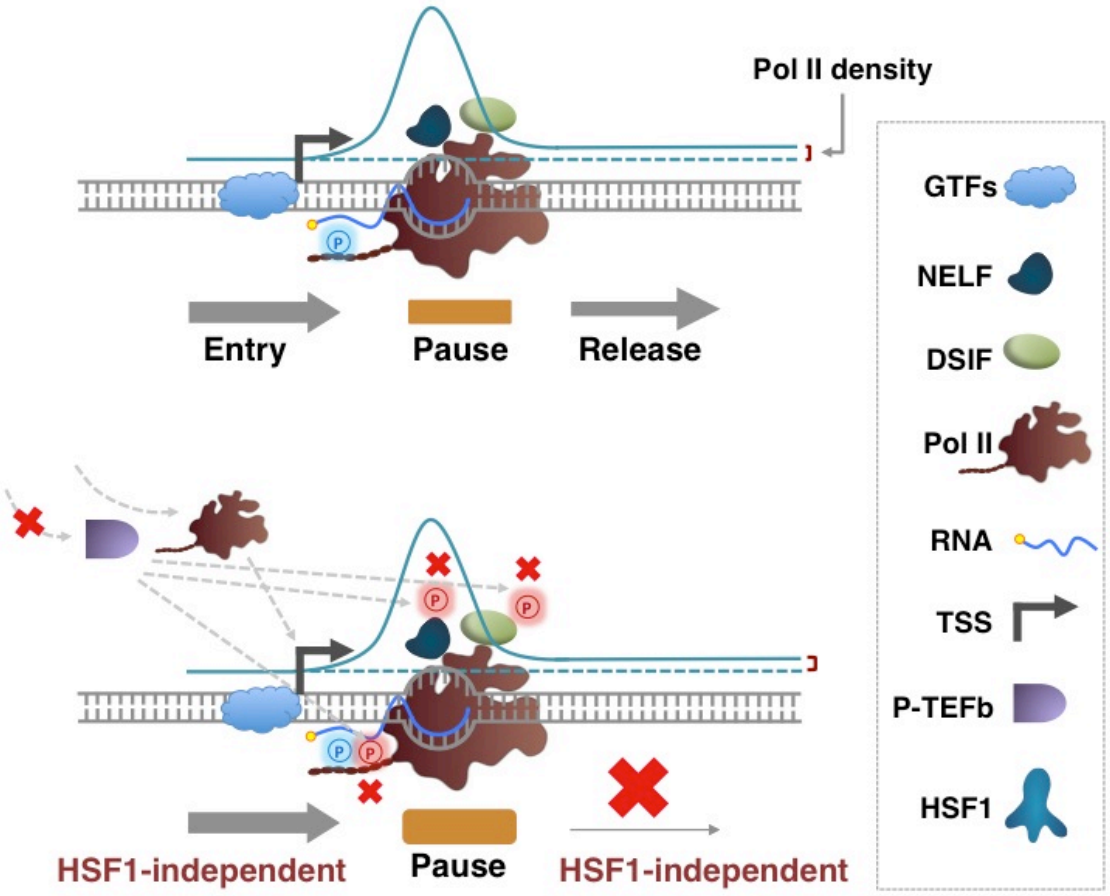


Figure 1.3. Mechanism of transcription repression during HSR
Global repression of transcription is caused by the inhibition of paused Pol II release from the promoter-proximal paused state. This inhibition and subsequent increase in paused Pol II is accompanied by decrease in P-TEFb occupancy at the promoters and it is independent of HSF1. The recruitment of Pol II is unaffected and independent of HSF1.

indicated that the P-TEFb binding at heat shock loci is dynamic (Zobeck et al., 2010), which supports the titration model in which HSF1 shuttles P-TEFb as it comes off the genes that will eventually be repressed. Alternatively, global P-TEFb may disengage at the onset of HSR requiring active recruitment by transcription factors such as HSF1 to the genes that are destined for upregulation. The failure to deliver P-TEFb to the remaining genes likely due to the lack of transcription factor binding results in their repression. It is also possible for P-TEFb to be limiting either by reduction in its kinase activity (Fujinaga et al., 1998), which is required for Pol II pause release, or due to its depletion from nucleus (Raychaudhuri et al., 2014) and likely sequestration in cytoplasmic stress granules.

Another interesting aspect of transcription regulation during HSR is the increase in paused Pol II density at both upregulated and downregulated genes. While it is theoretically possible that another transcription factor binds to all genes and recruits Pol II in HSF1-independent manner, the more plausible hypothesis is the mass action loading of Pol II that are liberated by global repression.

HSF1-regulated genes in stress

HSF1 regulated genes during stress are characterized in several model species (Brown et al., 2014; Duarte et al., 2016; Hahn et al., 2004; Mahat et al., 2016; Solís et al., 2016; Trinklein et al., 2004). The major constituents in

the repertoire of selectively induced genes are canonical *HSPs* that play a central role in chaperone machinery. But not all genes in the HSP family, as defined by a recent attempt to streamline the nomenclature of HSPs and auxiliary chaperone proteins (Kampinga et al., 2009), are stress-inducible at transcriptional level (Duarte et al., 2016; Mahat et al., 2016; Solís et al., 2016). While some of the uninduced HSPs could be induced in other type of stress, the general lack of HSF1 binding and HSE in their promoters indicate that they likely are not a transcriptional target of HSF1 (Mahat et al., 2016). Comparable expression of *HSPs* in wild type and *HSF1*$^{-/-}$ cells in unstressed condition indicates that the HSF1 does not play a role in basal expression of *HSPs* in mammalian cells. However, this observation also raises a critical question on the mechanism underlying basal expression of *HSPs*.

HSF1-regulated genes in absence of stress

The study of HSF1's role in unstressed cells is complicated by few factors. First, mammals have multiple paralogs of HSF1 and they bind to identical HSEs as homo- or hetero-trimers (Sandqvist et al., 2009). The lack of complete map of binding landscapes of different HSFs and their relative enrichment in different promoters confounds the contribution of each paralog in basal transcription regulation. Second, tools that can instantaneously and completely inhibit HSF1 function are yet to be developed. Conventional knockout (McMillan et al., 1998), CRISPR/Cas9 mediated genetic

manipulation (Solís et al., 2016), or RNAi based (Duarte et al., 2016) approach takes days to manifest their effect, which provides sufficient time for secondary effects to kick in. Tools that can instantaneously and almost completely inhibit HSF1 function such as inducible RNA aptamer (Salamanca et al., 2014) or small molecule inhibitors (Au et al., 2009; Whitesell and Lindquist, 2009; Yoon et al., 2011) are yet to be used to identify HSF1 regulated genes under basal conditions. A recent study using 'anchor-away' approach identified 18 genes that are constitutively regulated by HSF1 in yeast (Solís et al., 2016). However, CRISPR/Cas9 mediated genetic manipulation of HSF1 in the same study identified no genes under HSF1 control under basal conditions in mammals. There is a chance that this failure to find affected genes could have been a result of the cellular mechanisms to compensate for the loss of HSF1, because the CRISPR/Cas9 approach is not instantaneous. Moreover, previous studies have identified at least a few genes under HSF1's control in unstressed cells (Hayashida et al., 2010; Inouye et al., 2007). In the future, highly efficient tools that can be employed instantaneously, such as inducible RNA aptamers, potent small molecule inhibitors, or the mammalian 'anchor away' approach, can be used in conjunction with highly sensitive assays to decipher the transcriptional role of HSF1 in unstressed cells.

HSF1's role in cold shock

While much of the temperature-dependent stress is centered on a temperature increase, organisms are also exposed to cold environments. Seasonal cold acclimation trigger deep phenotypic transfiguration mediated by changes in global gene expression patterns and metabolome composition (Ragland et al., 2010). Rapid cold hardening (RCH) protects insects from cold injuries, which is primarily mediated by signaling cascade and protein phosphorylation (Kelty and Lee, 2001; Teets and Denlinger, 2013). Robust induction of *Hsp* genes mediated by HSF activation characterize both cases of long-term cold-acclimation and during recovery from cold shocks (Liu et al., 1994; Štětina et al., 2015). Understanding the mechanism of HSF1 in sensing a temperature decrease and the comparison of genome-wide transcriptional profiles in heat shock response and cold shock response can decipher the basic transcriptional program required for survival during stress.

HSF2 in HSR

Vertebrates have multiple paralogs of HSF1. HSF2 shares the highest degree of sequence homology (Ahn et al., 2001) and structural similarity (Jaeger et al., 2016; Neudegger et al., 2016) with HSF1. HSF1 and HSF2 are ubiquitously expressed unlike other HSFs (Akerfelt et al., 2010) and they bind to identical DNA sequences as homo- or hetero-trimers (Manuel et al., 2002; Sandqvist et al., 2009). This apparent redundancy has promoted the speculation of HSF2 being an insurance policy in evocation of the HSR. In

support of this hypothesis, HSF2 heterotrimerizes with HSF1 during HS and binds to the promoter of *Hsp70* gene (Ostling et al., 2007). It also binds to the promoters of selected *HSPs* constitutively and modulates their expression (Wilkerson et al., 2007). Surprisingly, human HSF2 but not HSF1 complements the viability defect associated with the deletion of yeast HSF gene (Liu et al., 1997), raising the debate whether HSF2 emerged by duplication of HSF1 or vice versa.

On the other hand, HSF2 plays distinct roles from HSF1 in development and non-stressed conditions. HSF2 is dispensable for *HSPs* induction during stress response (Jin et al., 2011; Kallio et al., 2002; Wang et al., 2003). A constitutive role of HSF2 is implicated in brain development (Kallio et al., 2002) and fetal alcohol syndrome (Fatimy et al., 2014). HSF2 is also known to play a role in diseases – it mitigates prostate cancer invasion (Björk et al., 2016) and mutations in HSF2 are associated with idiopathic azoospermia (Mou et al., 2013; Wilkerson et al., 2008). Proteotoxic stress induced by proteasome inhibition reduces the viability of both HSF1-knockout and HSF2-knockout MEFs but through different underlying mechanisms (Mou et al., 2013; Wilkerson et al., 2008).

The limited role of HSF2 during stress as inferred from a handful of studies on a few selected *HSPs* have defined the current view on HSF2 function during stress. In order to assess the genome-wide role of HSF2, our recent study examined the HSF2-dependent global transcriptional changes in

HSF1's presence or absence (Mahat et al., 2016). The lack of HSF2's role in stress-mediated induction of *HSPs* was confirmed by direct and quantitative measurement of transcription. Besides the complete reliance of *HSPs* and other chaperone assisting factors on HSF1 for their transcription induction during stress, the rest of the heat-induced genes were similarly regulated irrespective of HSF1, HSF2, or both. However, the pool of downregulated genes was significantly smaller in absence of HSF2 (Mahat et al., 2016), indicating the role of HSF2 in stress-mediated transcription repression. HSF2 is known to repress HSR in mitotic cells (Elsing et al., 2014), but the HSF2-dependent repression in freely cycling cells and the scale of repression was not observed before. The lack of HSF2 binding to the promoters of repressed genes (only ~4% of the HSF2-dependent repressed genes exhibit detectable HSF2 binding) indicate an indirect role of HSF2 in repression (unpublished data). One caveat however is that these observations were made using knockout cell lines, which could harbor compensatory or secondary effects. In order to discern the direct role of HSF2 in repression, instantaneous inhibition of HSF2 is required, likely through small molecule drugs or very rapid expression of domain-specific RNA aptamers.

HSR and HSF1 in therapeutics

Understanding the molecular events that occur immediately after exposure of cells to stress can unlock therapeutic potential in protein-folding

disorders and cancers. Cancer cells heavily depend on molecular chaperones to support imbalance in protein homeostasis (Dai et al., 2007) (Figure 1.4) and the increased expression and activity of chaperone machinery is directly controlled by HSF1 (Mendillo et al., 2012; Santagata et al., 2011; Scherz-Shouval et al., 2014). These findings have prompted the use of small molecule inhibitor of HSF1 as anti-cancer drug (Salamanca et al., 2014; Whitesell and Lindquist, 2009; Yoon et al., 2011). In contrast, neurological disorders such as Alzheimer, Parkinson, and Huntington, which are characterized by accumulation of protein-aggregates (Figure 1.4), have diminished molecular chaperone activity (Neef et al., 2011). Pharmaceutical companies are exploring therapeutic approaches to activate HSF1 in order to increase HSPs expression and enhance protein refolding (Neef et al., 2010b). This dichotomy in the roles of HSF1 in cancer and neurological diseases requires a deeper understanding of the precise molecular mechanism of HSF1-driven gene regulation before HSF1-based therapeutic tools can be clinically applied. Besides cancer and neurological disorders, HSF1 is implicated in aging and longevity (Morley and Morimoto, 2004), obesity (Ma et al., 2015), and cardiovascular diseases (Locke and Tanguay, 1996). Because of its widespread implication in biology and diseases, understanding the molecular mechanism of HSF1 function and identifying the genes regulated by HSF1 under different circumstances are more critical than ever.

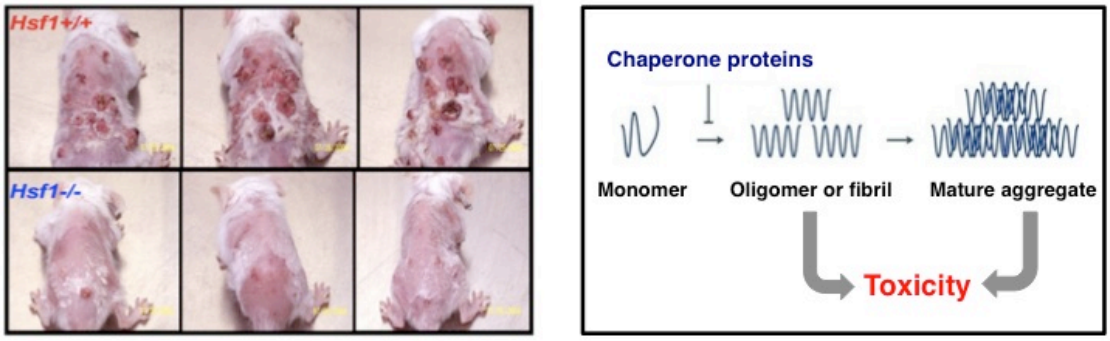


Figure 1.4. Opposing roles of HSF1 in cancer and neurological disorders
Cancer cells rely on HSF1 for tumorigenesis, prompting the use of small molecule inhibitor of HSF1 in cancer therapeutics. In contrast, Neurological disorders are characterized by diminished chaperone activity resulting in misfolded proteins, making pharmacological activation of HSF1 a promising therapeutics for neurological diseases. Adapted from (Dai et al., 2007; Neef et al., 2011).

REFERENCES

Abe, T., Yamamura, K., Gotoh, S., Kashimura, M., and Higashi, K. (1998). Concentration-dependent differential effects of N-acetyl-L-cysteine on the expression of HSP70 and metallothionein genes induced by cadmium in human amniotic cells. Biochim. Biophys. Acta *1380*, 123–132.

Abravaya, K., Myers, M.P., Murphy, S.P., and Morimoto, R.I. (1992). The human heat shock protein hsp70 interacts with HSF, the transcription factor that regulates heat shock gene expression. Genes & Development *6*, 1153–1164.

Adelman, K., and Lis, J.T. (2012). Promoter-proximal pausing of RNA polymerase II: emerging roles in metazoans. Nat. Rev. Genet. *13*, 720–731.

Ahn, S.G., Liu, P.C., Klyachko, K., Morimoto, R.I., and Thiele, D.J. (2001). The loop domain of heat shock transcription factor 1 dictates DNA-binding specificity and responses to heat stress. Genes & Development *15*, 2134–2145.

Ahn, S.-G., and Thiele, D.J. (2003). Redox regulation of mammalian heat shock factor 1 is essential for Hsp gene activation and protection from stress. Genes & Development *17*, 516–528.

Aït-Aïssa, S., Porcher, J., Arrigo, A., and Lambré, C. (2000). Activation of the hsp70 promoter by environmental inorganic and organic chemicals: relationships with cytotoxicity and lipophilicity. Toxicology *145*, 147–157.

Akerfelt, M., Morimoto, R.I., and Sistonen, L. (2010). Heat shock factors: integrators of cell stress, development and lifespan. Nat Rev Mol Cell Biol *11*, 545–555.

Ali, M.M.U., Roe, S.M., Vaughan, C.K., Meyer, P., Panaretou, B., Piper, P.W., Prodromou, C., and Pearl, L.H. (2006). Crystal structure of an Hsp90–nucleotide–p23/Sba1 closed chaperone complex. Nature *440*, 1013–1017.

Ananthan, J., Goldberg, A.L., and Voellmy, R. (1986). Abnormal proteins serve as eukaryotic stress signals and trigger the activation of heat shock genes. Science *232*, 522–524.

Andréasson, C., Fiaux, J., Rampelt, H., Druffel-Augustin, S., and Bukau, B. (2008). Insights into the structural dynamics of the Hsp110-Hsp70 interaction reveal the mechanism for nucleotide exchange activity. Proceedings of the National Academy of Sciences *105*, 16519–16524.

Au, Q., Zhang, Y., Barber, J.R., Ng, S.C., and Bin Zhang (2009). Identification of Inhibitors of HSF1 Functional Activity by High-Content Target-Based Screening. J Biomol Screen *14*, 1165–1175.

Björk, J.K., Åkerfelt, M., Joutsen, J., Puustinen, M.C., Cheng, F., Sistonen, L., and Nees, M. (2016). Heat-shock factor 2 is a suppressor of prostate cancer invasion. Oncogene *35*, 1770–1784.

Bolhassani, A., and Rafati, S. (2008). Heat-shock proteins as powerful weapons in vaccine development. Expert Rev Vaccines *7*, 1185–1199.

Boone, A.N., and Vijayan, M.M. (2002). Constitutive heat shock protein 70 (HSC70) expression in rainbow trout hepatocytes: effect of heat shock and heavy metal exposure. Comp. Biochem. Physiol. C Toxicol. Pharmacol. *132*, 223–233.

Brown, J.B., Boley, N., Eisman, R., May, G.E., Stoiber, M.H., Duff, M.O., Booth, B.W., Wen, J., Park, S., Suzuki, A.M., et al. (2014). Diversity and dynamics of the Drosophila transcriptome. Nature *512*, 393–399.

Budzyński, M.A., Puustinen, M.C., Joutsen, J., and Sistonen, L. (2015). Uncoupling Stress-Inducible Phosphorylation of Heat Shock Factor 1 from Its Activation. Mol. Cell. Biol. *35*, 2530–2540.

Chadli, A., Bouhouche, I., Sullivan, W., Stensgard, B., McMahon, N., Catelli, M.G., and Toft, D.O. (2000). Dimerization and N-terminal domain proximity

underlie the function of the molecular chaperone heat shock protein 90. Proc. Natl. Acad. Sci. U.S.a. *97*, 12524–12529.

Chang, J., Knowlton, A.A., Xu, F., and Wasser, J.S. (2001). Activation of the heat shock response: relationship to energy metabolites. A (31)P NMR study in rat hearts. Am. J. Physiol. Heart Circ. Physiol. *280*, H426–H433.

Clos, J., Westwood, J.T., Becker, P.B., Wilson, S., Lambert, K., and Wu, C. (1990). Molecular cloning and expression of a hexameric Drosophila heat shock factor subject to negative regulation. Cell *63*, 1085–1097.

Cotto, J.J., Kline, M., and Morimoto, R.I. (1996). Activation of heat shock factor 1 DNA binding precedes stress-induced serine phosphorylation. Evidence for a multistep pathway of regulation. J. Biol. Chem. *271*, 3355–3358.

Csermely, P., Kajtar, J., Hollosi, M., Jalsovszky, G., Holly, S., Kahn, C.R., Gergely, P., Soti, C., Mihály, K., and Somogyi, J. (1993). ATP induces a conformational change of the 90-kDa heat shock protein (hsp90). J. Biol. Chem. *268*, 1901–1907.

Dai, C., Whitesell, L., Rogers, A.B., and Lindquist, S. (2007). Heat shock factor 1 is a powerful multifaceted modifier of carcinogenesis. Cell *130*, 1005–1018.

De Maio, A., Santoro, M.G., Tanguay, R.M., and Hightower, L.E. (2012). Ferruccio Ritossa's scientific legacy 50 years after his discovery of the heat shock response: a new view of biology, a new society, and a new journal. Cell Stress Chaperones *17*, 139–143.

Deane, E.E., and Woo, N.Y.S. (2006). Impact of heavy metals and organochlorines on hsp70 and hsc70 gene expression in black sea bream fibroblasts. Aquat. Toxicol. *79*, 9–15.

Dobson, C.M., Šali, A., and Karplus, M. (1998). Protein Folding: A Perspective from Theory and Experiment. Angewandte Chemie International Edition *37*, 868–893.

Dollins, D.E., Warren, J.J., Immormino, R.M., and Gewirth, D.T. (2007). Structures of GRP94-Nucleotide Complexes Reveal Mechanistic Differences between the hsp90 Chaperones. Molecular Cell *28*, 41–56.

Dragovic, Z., Broadley, S.A., Shomura, Y., Bracher, A., and Hartl, F.U. (2006). Molecular chaperones of the Hsp110 family act as nucleotide exchange factors of Hsp70s. Embo J. *25*, 2519–2528.

Duarte, F.M., Fuda, N.J., Mahat, D.B., Core, L.J., Guertin, M.J., and Lis, J.T. (2016). Transcription factors GAF and HSF act at distinct regulatory steps to modulate stress-induced gene activation. Genes & Development *30*, 1731–1746.

Duina, A.A., Kalton, H.M., and Gaber, R.F. (1998). Requirement for Hsp90 and a CyP-40-type cyclophilin in negative regulation of the heat shock response. J. Biol. Chem. *273*, 18974–18978.

Easton, D.P., Kaneko, Y., and Subjeck, J.R. (2000). The hsp110 and Grp1 70 stress proteins: newly recognized relatives of the Hsp70s. Cell Stress Chaperones *5*, 276–290.

Elsing, A.N., Aspelin, C., Björk, J.K., Bergman, H.A., Himanen, S.V., Kallio, M.J., Roos-Mattjus, P., and Sistonen, L. (2014). Expression of HSF2 decreases in mitosis to enable stress-inducible transcription and cell survival. The Journal of Cell Biology *206*, 735–749.

Fatimy, El, R., Miozzo, F., Le Mouël, A., Abane, R., Schwendimann, L., Sabéran-Djoneidi, D., de Thonel, A., Massaoudi, I., Paslaru, L., Hashimoto-Torii, K., et al. (2014). Heat shock factor 2 is a stress-responsive mediator of neuronal migration defects in models of fetal alcohol syndrome. EMBO Mol Med *6*, 1043–1061.

Fehrenbach, E., and Niess, A.M. (1999). Role of heat shock proteins in the exercise response. Exerc Immunol Rev *5*, 57–77.

Fuda, N.J., Ardehali, M.B., and Lis, J.T. (2009). Defining mechanisms that regulate RNA polymerase II transcription in vivo. Nature *461*, 186–192.

Fujimoto, M., Takaki, E., Takii, R., Tan, K., Prakasam, R., Hayashida, N., Iemura, S.-I., Natsume, T., and Nakai, A. (2012). RPA Assists HSF1 Access to Nucleosomal DNA by Recruiting Histone Chaperone FACT. Molecular Cell *48*, 182–194.

Fujinaga, K., Cujec, T.P., Peng, J., and Garriga, J. (1998). The ability of positive transcription elongation factor B to transactivate human immunodeficiency virus transcription depends on a functional kinase domain, cyclin T1, …. Journal of ….

Gallo, G.J., Prentice, H., and Kingston, R.E. (1993). Heat shock factor is required for growth at normal temperatures in the fission yeast Schizosaccharomyces pombe. Mol. Cell. Biol. *13*, 749–761.

Genevaux, P., Georgopoulos, C., and Kelley, W.L. (2007). The Hsp70 chaperone machines of Escherichia coli: a paradigm for the repartition of chaperone functions. Molecular Microbiology *66*, 840–857.

Giardina, C., and Lis, J.T. (1995a). Dynamic protein-DNA architecture of a yeast heat shock promoter. Mol. Cell. Biol. *15*, 2737–2744.

Giardina, C., and Lis, J.T. (1995b). Sodium Salicylate and Yeast Heat Shock Gene Transcription. J. Biol. Chem. *270*, 10369–10372.

Goodson, M.L., and Sarge, K.D. (1995). Heat-inducible DNA Binding of Purified Heat Shock Transcription Factor 1. J. Biol. Chem. *270*, 2447–2450.

Grenert, J.P., Sullivan, W.P., Fadden, P., Haystead, T.A.J., Clark, J., Mimnaugh, E., Krutzsch, H., Ochel, H.-J., Schulte, T.W., Sausville, E., et al. (1997). The Amino-terminal Domain of Heat Shock Protein 90 (hsp90) That Binds Geldanamycin Is an ATP/ADP Switch Domain That Regulates hsp90 Conformation. J. Biol. Chem. *272*, 23843–23850.

Guertin, M.J., and Lis, J.T. (2010). Chromatin landscape dictates HSF binding to target DNA elements. PLoS Genet. *6*, e1001114.

Guettouche, T., Boellmann, F., Lane, W.S., and Voellmy, R. (2005). Analysis of phosphorylation of human heat shock factor 1 in cells experiencing a stress. BMC Biochemistry 2005 6:1 *6*, 4.

Guo, Y., Guettouche, T., Fenna, M., Boellmann, F., Pratt, W.B., Toft, D.O., Smith, D.F., and Voellmy, R. (2001). Evidence for a mechanism of repression of heat shock factor 1 transcriptional activity by a multichaperone complex. J. Biol. Chem. *276*, 45791–45799.

Hahn, J.-S., Hu, Z., Thiele, D.J., and Iyer, V.R. (2004). Genome-wide analysis of the biology of stress responses through heat shock transcription factor. Mol. Cell. Biol. *24*, 5249–5256.

Harrison, C.J., Bohn, A., and Nelson, H.C.M. (1994). Crystal Structure of the DNA Binding Domain of Heat Shock Transcription Factor. Science *263*.

Hayashida, N., Fujimoto, M., Tan, K., Prakasam, R., Shinkawa, T., Li, L., Ichikawa, H., Takii, R., and Nakai, A. (2010). Heat shock factor 1 ameliorates proteotoxicity in cooperation with the transcription factor NFAT. Embo J. *29*, 3459–3469.

Hensler, T., Köller, M., Alouf, J.E., and König, W. (1991). Bacterial toxins induce heat shock proteins in human neutrophils. Biochem. Biophys. Res. Commun. *179*, 872–879.

Hentze, N., Le Breton, L., Wiesner, J., Kempf, G., and Mayer, M.P. (2016). Molecular mechanism of thermosensory function of human heat shock transcription factor Hsf1. Elife *5*, 1153.

Herbst, R., Schäfer, U., and Seckler, R. (1997). Equilibrium Intermediates in the Reversible Unfolding of Firefly (Photinus pyralis) Luciferase. J. Biol. Chem. *272*, 7099–7105.

Hietakangas, V., Ahlskog, J.K., Jakobsson, A.M., Hellesuo, M., Sahlberg, N.M., Holmberg, C.I., Mikhailov, A., Palvimo, J.J., Pirkkala, L., and Sistonen, L. (2003). Phosphorylation of serine 303 is a prerequisite for the stress-inducible SUMO modification of heat shock factor 1. Mol. Cell. Biol. *23*, 2953–2968.

Holmberg, C.I., Hietakangas, V., Mikhailov, A., Rantanen, J.O., Kallio, M., Meinander, A., Hellman, J., Morrice, N., MacKintosh, C., Morimoto, R.I., et al. (2001). Phosphorylation of serine 230 promotes inducible transcriptional activity of heat shock factor 1. Embo J. *20*, 3800–3810.

Hong, S., Kim, S.H., Heo, M.A., Choi, Y.H., Park, M.J., Yoo, M.A., Kim, H.D., Kang, H.S., and Cheong, J. (2004). Coactivator ASC-2 mediates heat shock factor 1-mediated transactivation dependent on heat shock. FEBS Letters *559*, 165–170.

Hosaka, S., Nakatsura, T., Tsukamoto, H., Hatayama, T., Baba, H., and Nishimura, Y. (2006). Synthetic small interfering RNA targeting heat shock protein 105 induces apoptosis of various cancer cells both in vitro and in vivo. Cancer Science *97*, 623–632.

Høj, A., and Jakobsen, B.K. (1994). A short element required for turning off heat shock transcription factor: evidence that phosphorylation enhances deactivation. Embo J. *13*, 2617–2624.

Inouye, S., Fujimoto, M., Nakamura, T., Takaki, E., Hayashida, N., Hai, T., and Nakai, A. (2007). Heat shock transcription factor 1 opens chromatin structure of interleukin-6 promoter to facilitate binding of an activator or a repressor. J. Biol. Chem. *282*, 33210–33217.

Jaeger, A.M., Pemble, C.W., Sistonen, L., and Thiele, D.J. (2016). Structures of HSF2 reveal mechanisms for differential regulation of human heat-shock factors. Nat. Struct. Mol. Biol. *23*, 147–154.

Jakobsen, B.K., and Pelham, H.R. (1991). A conserved heptapeptide restrains the activity of the yeast heat shock transcription factor. Embo J. *10*, 369–375.

Jarosz, D.F., and Lindquist, S. (2010). Hsp90 and environmental stress transform the adaptive value of natural genetic variation. Science *330*, 1820–1824.

Jedlicka, P., Mortin, M.A., and Wu, C. (1997). Multiple functions of Drosophilaheat shock transcription factor in vivo. Embo J. *16*, 2452–2462.

Jin, X., Eroglu, B., Moskophidis, D., and Mivechi, N.F. (2011). Targeted deletion of Hsf1, 2, and 4 genes in mice. Methods Mol. Biol. *787*, 1–20.

Jurivich, D.A., Pachetti, C., Qiu, L., and Welk, J.F. (1995). Salicylate triggers heat shock factor differently than heat. J. Biol. Chem. *270*, 24489–24495.

Jurivich, D.A., Sistonen, L., Kroes, R.A., and Morimoto, R.I. (1992). Effect of sodium salicylate on the human heat shock response. Science *255*.

Kakkar, V., Meister-Broekema, M., Minoia, M., Carra, S., and Kampinga, H.H. (2014). Barcoding heat shock proteins to human diseases: looking beyond the heat shock response. Dis Model Mech *7*, 421–434.

Kallio, M., Chang, Y., Manuel, M., Alastalo, T.-P., Rallu, M., Gitton, Y., Pirkkala, L., Loones, M.-T., Paslaru, L., Larney, S., et al. (2002). Brain abnormalities, defective meiotic chromosome synapsis and female subfertility in HSF2 null mice. Embo J. *21*, 2591–2601.

Kampinga, H.H., Hageman, J., Vos, M.J., Kubota, H., Tanguay, R.M., Bruford, E.A., Cheetham, M.E., Chen, B., and Hightower, L.E. (2009). Guidelines for the nomenclature of the human heat shock proteins. Cell Stress Chaperones *14*, 105–111.

Kang, G.-Y., Kim, E.-H., Lee, H.-J., Gil, N.-Y., Cha, H.-J., and Lee, Y.-S. (2015). Heat shock factor 1, an inhibitor of non-homologous end joining repair. Oncotarget *6*, 29712–29724.

Kantengwa, S., and Polla, B.S. (1993). Phagocytosis of Staphylococcus aureus induces a selective stress response in human monocytes-macrophages (M phi): modulation by M phi differentiation and by iron. Infect. Immun. *61*, 1281–1287.

Kelty, J.D., and Lee, R.E. (2001). Rapid cold-hardening of Drosophila melanogaster (Diptera: Drosophiladae) during ecologically based thermoperiodic cycles. J. Exp. Biol. *204*, 1659–1666.

Kim, S., and Gross, D.S. (2013). Mediator recruitment to heat shock genes requires dual Hsf1 activation domains and mediator tail subunits Med15 and Med16. Journal of Biological Chemistry *288*, 12197–12213.

Kline, M.P., and Morimoto, R.I. (1997). Repression of the heat shock factor 1 transcriptional activation domain is modulated by constitutive phosphorylation. Mol. Cell. Biol. *17*, 2107–2115.

Knauf, U., Newton, E.M., Kyriakis, J., and Kingston, R.E. (1996). Repression of human heat shock factor 1 activity at control temperature by phosphorylation. Genes & Development *10*, 2782–2793.

Kubelka, J., Hofrichter, J., and Eaton, W.A. (2004). The protein folding “speed limit.” Current Opinion in Structural Biology *14*, 76–88.

Kugel, J.F., and Goodrich, J.A. (2006). Beating the Heat: A Translation Factor and an RNA Mobilize the Heat Shock Transcription Factor HSF1. Molecular Cell *22*, 153–154.

Lee, Y.-J., Kim, E.-H., Lee, J.S., Jeoung, D., Bae, S., Kwon, S.H., and Lee, Y.-S. (2008). HSF1 as a Mitotic Regulator: Phosphorylation of HSF1 by Plk1 Is Essential for Mitotic Progression. Cancer Res. *68*, 7550–7560.

Leon, L.R. (2016). Common mechanisms for the adaptive responses to exercise and heat stress. J. Appl. Physiol. *120*, 662–663.

Li, D., Marchenko, N.D., Schulz, R., and Fischer, V. (2011). Functional inactivation of endogenous MDM2 and CHIP by HSP90 causes aberrant stabilization of mutant p53 in human cancer cells (Molecular Cancer …).

Lin, H., Opler, M., Head, M., Blank, M., and Goodman, R. (1997). Electromagnetic field exposure induces rapid, transitory heat shock factor activation in human cells. J. Cell. Biochem. *66*, 482–488.

Lindquist, S. (1986). The heat-shock response. Annu. Rev. Biochem. *55*, 1151–1191.

Lindquist, S., and Kim, G. (1996). Heat-shock protein 104 expression is sufficient for thermotolerance in yeast. Proc. Natl. Acad. Sci. U.S.a. *93*, 5301–5306.

Lis, J.T., Mason, P., Peng, J., Price, D.H., and Werner, J. (2000). P-TEFb kinase recruitment and function at heat shock loci. Genes & Development *14*, 792–803.

Lis, J.T., Simon, J.A., and Sutton, C.A. (1983). New heat shock puffs and β=galactosidase activity resulting from transformation of Drosophila with an hsp70-lacZ hybrid gene. Cell *35*, 403–410.

Littlefield, O., and Nelson, H.C.M. (1999). A new use for the 'wing' of the “winged” helix-turn-helix motif in the HSFl[ndash]lDNA cocrystal. Nat. Struct. Mol. Biol. *6*, 464–470.

Liu, A.Y., Bian, H., Huang, L.E., and Lee, Y.K. (1994). Transient cold shock induces the heat shock response upon recovery at 37 degrees C in human cells. J. Biol. Chem. *269*, 14768–14775.

Liu, G., Xiao, R., Ciccosanti, C., Janjua, H., Acton, T.B., Lee, H., Wang, H.B., Huang, Y.B., Everett, J.K., Montelione, G.T., et al. (2011). Solution NMR Structure of Heat shock factor protein 1 DNA binding domain from homo sapiens, Northeast Structural Genomics Consortium Target HR3023C.

Liu, J., Squibb, K.S., Akkerman, M., Nordberg, G.F., Lipsky, M., and Fowler, B.A. (1996). Cytotoxicity, zinc protection, and stress protein induction in rat proximal tubule cells exposed to cadmium chloride in primary cell culture. Ren Fail *18*, 867–882.

Liu, X.D., Liu, P.C., Santoro, N., and Thiele, D.J. (1997). Conservation of a stress response: human heat shock transcription factors functionally substitute for yeast HSF. Embo J. *16*, 6466–6477.

Locke, M., and Tanguay, R.M. (1996). Diminished heat shock response in the aged myocardium. Cell Stress Chaperones *1*, 251–260.

Löw-Friedrich, I., and Schoeppe, W. (1991). Effects of Calcium Channel Blockers on Stress Protein Synthesis in Cardiac Myocytes. Journal of Cardiovascular Pharmacology *17*, 800.

Luft, J.C., Benjamin, I.J., Mestril, R., and Dix, D.J. (2001). Heat shock factor 1–mediated thermotolerance prevents cell death and results in G2/M cell cycle arrest. Cell Stress & Chaperones *6*, 326-336

Ma, X., Xu, L., Alberobello, A.T., Gavrilova, O., Bagattin, A., Skarulis, M., Liu, J., Finkel, T., and Mueller, E. (2015). Celastrol Protects against Obesity and Metabolic Dysfunction through Activation of a HSF1-PGC1α Transcriptional Axis. Cell Metab. *22*, 695–708.

Mahat, D.B., and Lis, J.T. (2016). Use of conditioned media is critical for studies of regulation in response to rapid heat shock. Cell Stress Chaperones 1–8.

Mahat, D.B., Salamanca, H.H., Duarte, F.M., Danko, C.G., and Lis, J.T. (2016). Mammalian Heat Shock Response and Mechanisms Underlying Its Genome-wide Transcriptional Regulation. Molecular Cell *62*, 63–78.

Manuel, M., Rallu, M., Loones, M.-T., Zimarino, V., Mezger, V., and Morange, M. (2002). Determination of the consensus binding sequence for the purified

embryonic heat shock factor 2. Eur. J. Biochem. *269*, 2527–2537.

Mattoo, R.U.H., Sharma, S.K., Priya, S., Finka, A., and Goloubinoff, P. (2013). Hsp110 is a bona fide chaperone using ATP to unfold stable misfolded polypeptides and reciprocally collaborate with Hsp70 to solubilize protein aggregates. Journal of Biological Chemistry *288*, 21399–21411.

McMillan, D.R., Xiao, X., Shao, L., Graves, K., and Benjamin, I.J. (1998). Targeted disruption of heat shock transcription factor 1 abolishes thermotolerance and protection against heat-inducible apoptosis. J. Biol. Chem. *273*, 7523–7528.

Mendillo, M.L., Santagata, S., Koeva, M., Bell, G.W., Hu, R., Tamimi, R.M., Fraenkel, E., Ince, T.A., Whitesell, L., and Lindquist, S. (2012). HSF1 Drives a Transcriptional Program Distinct from Heat Shock to Support Highly Malignant Human Cancers. Cell *150*, 549–562.

Meyer, K.D., Patil, D.P., Zhou, J., Zinoviev, A., Skabkin, M.A., Elemento, O., Pestova, T.V., Qian, S.-B., and Jaffrey, S.R. (2015). 5′ UTR m6A Promotes Cap-Independent Translation. Cell *163*, 999–1010.

Moraitis, C., and Curran, B.P.G. (2004). Reactive oxygen species may influence the heat shock response and stress tolerance in the yeast Saccharomyces cerevisiae. Yeast *21*, 313–323.

Morimoto, R.I. (1998a). Regulation of the heat shock transcriptional response: cross talk between a family of heat shock factors, molecular chaperones, and negative regulators. Genes & Development *12*, 3788–3796.

Morimoto, R.I. (1998b). Regulation of the heat shock transcriptional response: cross talk between a family of heat shock factors, molecular chaperones, and negative regulators. Genes & Development *12*, 3788–3796.

Morley, J.F., and Morimoto, R.I. (2004). Regulation of longevity in Caenorhabditis elegans by heat shock factor and molecular chaperones. Mol.

Biol. Cell *15*, 657–664.

Mosser, D.D., Kotzbauer, P.T., Sarge, K.D., and Morimoto, R.I. (1990). In vitro activation of heat shock transcription factor DNA-binding by calcium and biochemical conditions that affect protein conformation. Proc. Natl. Acad. Sci. U.S.a. *87*, 3748–3752.

Mou, L., Wang, Y., Li, H., Huang, Y., Jiang, T., Huang, W., Li, Z., Chen, J., Xie, J., Liu, Y., et al. (2013). A dominant-negative mutation of HSF2 associated with idiopathic azoospermia. Hum. Genet. *132*, 159–165.

Murshid, A., Gong, J., and Calderwood, S.K. (2008). Heat-shock proteins in cancer vaccines: agents of antigen cross-presentation. Expert Rev Vaccines *7*, 1019–1030.

Nadeau, K., Das, A., and Walsh, C.T. (1993). Hsp90 chaperonins possess ATPase activity and bind heat shock transcription factors and peptidyl prolyl isomerases. J. Biol. Chem. *268*, 1479–1487.

Neckers, L., and Workman, P. (2012). Hsp90 molecular chaperone inhibitors: are we there yet? Clin. Cancer Res. *18*, 64–76.

Neef, D.W., Jaeger, A.M., and Thiele, D.J. (2011). Heat shock transcription factor 1 as a therapeutic target in neurodegenerative diseases. Nat Rev Drug Discov *10*, 930–944.

Neef, D.W., Jaeger, A.M., Gomez-Pastor, R., Willmund, F., Frydman, J., and Thiele, D.J. (2014). A direct regulatory interaction between chaperonin TRiC and stress-responsive transcription factor HSF1. Cell Rep *9*, 955–966.

Neef, D.W., Turski, M.L., and Thiele, D.J. (2010a). Modulation of heat shock transcription factor 1 as a therapeutic target for small molecule intervention in neurodegenerative disease. PLoS Biol. *8*, e1000291.

Neef, D.W., Turski, M.L., and Thiele, D.J. (2010b). Modulation of Heat Shock Transcription Factor 1 as a Therapeutic Target for Small Molecule Intervention in Neurodegenerative Disease. PLoS Biol. *8*, e1000291.

Neudegger, T., Verghese, J., Hayer-Hartl, M., Hartl, F.U., and Bracher, A. (2016). Structure of human heat-shock transcription factor 1 in complex with DNA. Nat. Struct. Mol. Biol. *23*, 140–146.

Ohama, N., Kusakabe, K., Mizoi, J., Zhao, H., Kidokoro, S., Koizumi, S., Takahashi, F., Ishida, T., Yanagisawa, S., Shinozaki, K., et al. (2016). The Transcriptional Cascade in the Heat Stress Response of Arabidopsis Is Strictly Regulated at the Level of Transcription Factor Expression. The Plant Cell Online *28*, 181–201.

Ostling, P., Björk, J.K., Roos-Mattjus, P., Mezger, V., and Sistonen, L. (2007). Heat shock factor 2 (HSF2) contributes to inducible expression of hsp genes through interplay with HSF1. J. Biol. Chem. *282*, 7077–7086.

Park, J.M., Werner, J., Kim, J.M., Lis, J.T., and Kim, Y.-J. (2001). Mediator, Not Holoenzyme, Is Directly Recruited to the Heat Shock Promoter by HSF upon Heat Shock. Molecular Cell *8*, 9–19.

Parker, C.S., and Topol, J. (1984). A Drosophila RNA polymerase II transcription factor binds to the regulatory site of an hsp 70 gene. Cell *37*, 273–283.

Perisic, O., Xiao, H., and Lis, J.T. (1989). Stable binding of Drosophila heat shock factor to head-to-head and tail-to-tail repeats of a conserved 5 bp recognition unit. Cell *59*, 797–806.

Petesch, S.J., and Lis, J.T. (2008). Rapid, transcription-independent loss of nucleosomes over a large chromatin domain at Hsp70 loci. Cell *134*, 74–84.

Petesch, S.J., and Lis, J.T. (2012). Activator-induced spread of poly(ADP-ribose) polymerase promotes nucleosome loss at Hsp70. Molecular Cell *45*,

64–74.

Prahlad, V., Cornelius, T., and Morimoto, R.I. (2008). Regulation of the Cellular Heat Shock Response in Caenorhabditis elegans by Thermosensory Neurons. Science *320*, 811–814.

Rabindran, S.K., Haroun, R.I., Clos, J., Wisniewski, J., and Wu, C. (1993). Regulation of heat shock factor trimer formation: role of a conserved leucine zipper. Science *259*, 230–234.

Ragland, G.J., Denlinger, D.L., and Hahn, D.A. (2010). Mechanisms of suspended animation are revealed by transcript profiling of diapause in the flesh fly. Proceedings of the National Academy of Sciences *107*, 14909–14914.

Ramaglia, V., Harapa, G.M., White, N., and Buck, L.T. (2004). Bacterial infection and tissue-specific Hsp72, -73 and -90 expression in western painted turtles. Comp. Biochem. Physiol. C Toxicol. Pharmacol. *138*, 139–148.

Raviol, H., Sadlish, H., Rodriguez, F., Mayer, M.P., and Bukau, B. (2006). Chaperone network in the yeast cytosol: Hsp110 is revealed as an Hsp70 nucleotide exchange factor. Embo J. *25*, 2510–2518.

Raychaudhuri, S., Loew, C., Körner, R., Pinkert, S., Theis, M., Hayer-Hartl, M., Buchholz, F., and Hartl, F.U. (2014). Interplay of Acetyltransferase EP300 and the Proteasome System in Regulating Heat Shock Transcription Factor 1. Cell *156*, 975–985.

Richter, K., Haslbeck, M., and Buchner, J. (2010). The heat shock response: life on the verge of death. Molecular Cell *40*, 253–266.

Ritossa, F. (1962). A new puffing pattern induced by temperature shock and DNP in drosophila. Experientia *18*, 571–573.

Roccheri, M.C., Agnello, M., Bonaventura, R., and Matranga, V. (2004). Cadmium induces the expression of specific stress proteins in sea urchin embryos. Biochem. Biophys. Res. Commun. *321*, 80–87.

Rokavec, M., Wu, W., and Luo, J.-L. (2012). IL6-Mediated Suppression of miR-200c Directs Constitutive Activation of Inflammatory Signaling Circuit Driving Transformation and Tumorigenesis. Molecular Cell *45*, 777–789.

Rougvie, A.E., and Lis, J.T. (1988). The RNA polymerase II molecule at the 5' end of the uninduced hsp70 gene of D. melanogaster is transcriptionally engaged. Cell *54*, 795–804.

Rutherford, S.L., and Lindquist, S. (1998). Hsp90 as a capacitor for morphological evolution. Nature *396*, 336–342.

Rüdiger, S., Buchberger, A., and Bukau, B. (1997). Interaction of Hsp70 chaperones with substrates. Nat. Struct. Mol. Biol. *4*, 342–349.

Salamanca, H.H., Antonyak, M.A., Cerione, R.A., Shi, H., and Lis, J.T. (2014). Inhibiting heat shock factor 1 in human cancer cells with a potent RNA aptamer. PLoS ONE *9*, e96330.

Sanchez, Y., and Lindquist, S.L. (1990). HSP104 required for induced thermotolerance. Science.

Sandqvist, A., Björk, J.K., Akerfelt, M., Chitikova, Z., Grichine, A., Vourc'h, C., Jolly, C., Salminen, T.A., Nymalm, Y., and Sistonen, L. (2009). Heterotrimerization of heat-shock factors 1 and 2 provides a transcriptional switch in response to distinct stimuli. Mol. Biol. Cell *20*, 1340–1347.

Santagata, S., Hu, R., Lin, N.U., Mendillo, M.L., Collins, L.C., Hankinson, S.E., Schnitt, S.J., Whitesell, L., Tamimi, R.M., Lindquist, S., et al. (2011). High levels of nuclear heat-shock factor 1 (HSF1) are associated with poor prognosis in breast cancer. Proceedings of the National Academy of Sciences *108*, 18378–18383.

Santagata, S., Mendillo, M.L., Tang, Y.-C., Subramanian, A., Perley, C.C., Roche, S.P., Wong, B., Narayan, R., Kwon, H., Koeva, M., et al. (2013). Tight coordination of protein translation and HSF1 activation supports the anabolic malignant state. Science *341*, 1238303–1238303.

Scherz-Shouval, R., Santagata, S., Mendillo, M.L., Sholl, L.M., Ben-Aharon, I., Beck, A.H., Dias-Santagata, D., Koeva, M., Stemmer, S.M., Whitesell, L., et al. (2014). The Reprogramming of Tumor Stroma by HSF1 Is a Potent Enabler of Malignancy. Cell *158*, 564–578.

Shalgi, R., Hurt, J.A., Krykbaeva, I., Taipale, M., Lindquist, S., and Burge, C.B. (2013). Widespread regulation of translation by elongation pausing in heat shock. Molecular Cell *49*, 439–452.

Shalgi, R., Hurt, J.A., Lindquist, S., and Burge, C.B. (2014). Widespread Inhibition of Posttranscriptional Splicing Shapes the Cellular Transcriptome following Heat Shock. Cell Rep *7*, 1362–1370.

Shamovsky, I., Ivannikov, M., Kandel, E.S., Gershon, D., and Nudler, E. (2006). RNA-mediated response to heat shock in mammalian cells. Nature *440*, 556–560.

Shiau, A.K., Harris, S.F., Southworth, D.R., and Agard, D.A. (2006). Structural Analysis of E. coli hsp90 Reveals Dramatic Nucleotide-Dependent Conformational Rearrangements. Cell *127*, 329–340.

Singer, C., Zimmermann, S., and Sures, B. (2005). Induction of heat shock proteins (hsp70) in the zebra mussel (Dreissena polymorpha) following exposure to platinum group metals (platinum, palladium and rhodium): comparison with lead and cadmium exposures. Aquat. Toxicol. *75*, 65–75.

Smith, S.T., Petruk, S., Sedkov, Y., Cho, E., Tillib, S., Canaani, E., and Mazo, A. (2004). Modulation of heat shock gene expression by the TAC1 chromatin-modifying complex. Nat. Cell Biol. *6*, 162–167.

Solís, E.J., Pandey, J.P., Zheng, X., Jin, D.X., Gupta, P.B., Airoldi, E.M., Pincus, D., and Denic, V. (2016). Defining the Essential Function of Yeast Hsf1 Reveals a Compact Transcriptional Program for Maintaining Eukaryotic Proteostasis. Molecular Cell *63*, 60–71.

Soncin, F., Zhang, X., Chu, B., Wang, X., Asea, A., Ann Stevenson, M., Sacks, D.B., and Calderwood, S.K. (2003). Transcriptional activity and DNA binding of heat shock factor-1 involve phosphorylation on threonine 142 by CK2. Biochem. Biophys. Res. Commun. *303*, 700–706.

Sorger, P.K., and Pelham, H.R. (1988). Yeast heat shock factor is an essential DNA-binding protein that exhibits temperature-dependent phosphorylation. Cell *54*, 855–864.

Štětina, T., Koštál, V., and Korbelová, J. (2015). The Role of Inducible Hsp70, and Other Heat Shock Proteins, in Adaptive Complex of Cold Tolerance of the Fruit Fly (Drosophila melanogaster). PLoS ONE *10*, e0128976.

Taipale, M., Jarosz, D.F., and Lindquist, S. (2010). HSP90 at the hub of protein homeostasis: emerging mechanistic insights. Nat Rev Mol Cell Biol *11*, 515–528.

Teets, N.M., and Denlinger, D.L. (2013). Physiological mechanisms of seasonal and rapid cold-hardening in insects. Physiological Entomology *38*, 105–116.

Tissiéres, A., Mitchell, H.K., and Tracy, U.M. (1974). Protein synthesis in salivary glands of Drosophila melanogaster: Relation to chromosome puffs. Journal of Molecular Biology *84*, 389IN9393–392IN12398.

Torres, F.A., and Bonner, J.J. (1995). Genetic identification of the site of DNA contact in the yeast heat shock transcription factor. Mol. Cell. Biol. *15*, 5063–5070.

Trautinger, F., Kindås-Mügge, I., Knobler, R.M., and Hönigsmann, H. (1996).

Stress proteins in the cellular response to ultraviolet radiation. Journal of Photochemistry and Photobiology B: Biology *35*, 141–148.

Trepel, J., Mollapour, M., Giaccone, G., and Neckers, L. (2010). Targeting the dynamic HSP90 complex in cancer. Nat. Rev. Cancer *10*, 537–549.

Trinklein, N.D., Murray, J.I., Hartman, S.J., Botstein, D., and Myers, R.M. (2004). The role of heat shock transcription factor 1 in the genome-wide regulation of the mammalian heat shock response. Mol. Biol. Cell *15*, 1254–1261.

Trott, A., West, J.D., Klaić, L., Westerheide, S.D., Silverman, R.B., Morimoto, R.I., and Morano, K.A. (2008). Activation of heat shock and antioxidant responses by the natural product celastrol: transcriptional signatures of a thiol-targeted molecule. Mol. Biol. Cell *19*, 1104–1112.

Verghese, J., Abrams, J., Wang, Y., and Morano, K.A. (2012). Biology of the heat shock response and protein chaperones: budding yeast (Saccharomyces cerevisiae) as a model system. Microbiol. Mol. Biol. Rev. *76*, 115–158.

Vuister, G.W., Kim, S.J., Orosz, A., and Marquardt, J. (1994). Solution structure of the DNA-binding domain of Drosophila heat shock transcription factor. Nature Structural ….

Wallen, E.S., Buettner, G.R., and Moseley, P.L. (1997). Oxidants differentially regulate the heat shock response. Int J Hyperthermia *13*, 517–524.

Wang, G., Zhang, J., Moskophidis, D., and Mivechi, N.F. (2003). Targeted disruption of the heat shock transcription factor (hsf)-2 gene results in increased embryonic lethality, neuronal defects, and reduced spermatogenesis. Genesis *36*, 48–61.

Warchałowska-Sliwa, E., Niklińska, M., Görlich, A., Michailova, P., and Pyza, E. (2005). Heavy metal accumulation, heat shock protein expression and cytogenetic changes in Tetrix tenuicornis (L.) (Tetrigidae, Orthoptera) from

polluted areas. Environ. Pollut. *133*, 373–381.

Westerheide, S.D., Anckar, J., Stevens, S.M., Sistonen, L., and Morimoto, R.I. (2009). Stress-inducible regulation of heat shock factor 1 by the deacetylase SIRT1. Science *323*, 1063–1066.

Westwood, J.T., Clos, J., and Wu, C. (1991). Stress-induced oligomerization and chromosomal relocalization of heat-shock factor. Nature *353*, 822–827.

Whitesell, L., and Lindquist, S. (2009). Inhibiting the transcription factor HSF1 as an anticancer strategy. Expert Opin. Ther. Targets *13*, 469–478.

Wiederrecht, G., Seto, D., and Parker, C.S. (1988). Isolation of the gene encoding the S. cerevisiae heat shock transcription factor. Cell *54*, 841–853.

Wilkerson, D.C., Murphy, L.A., and Sarge, K.D. (2008). Interaction of HSF1 and HSF2 with the Hspa1b promoter in mouse epididymal spermatozoa. Biol. Reprod. *79*, 283–288.

Wilkerson, D.C., Skaggs, H.S., and Sarge, K.D. (2007). HSF2 binds to the Hsp90, Hsp27, and c-Fos promoters constitutively and modulates their expression. Cell Stress Chaperones *12*, 283–290.

Wirth, D., Christians, E., Munaut, C., Dessy, C., Foidart, J.-M., and Gustin, P. (2002). Differential heat shock gene hsp70-1 response to toxicants revealed by in vivo study of lungs in transgenic mice. Cell Stress Chaperones *7*, 387–395.

Wu, C. (1995). Heat shock transcription factors: structure and regulation. Annu. Rev. Cell Dev. Biol. *11*, 441–469.

Xiao, H., and Lis, J.T. (1988). Germline transformation used to define key features of heat-shock response elements. Science *239*, 1139–1142.

Yamada, P., Amorim, F., Moseley, P., and Schneider, S. (2008). Heat shock protein 72 response to exercise in humans. Sports Med *38*, 715–733.

Yamagishi, N., Saito, Y., and Hatayama, T. (2008). Mammalian 105 kDa heat shock family proteins suppress hydrogen peroxide-induced apoptosis through a p38 MAPK-dependent mitochondrial pathway in HeLa cells. Febs J. *275*, 4558–4570.

Yamagishi, N., Yokota, M., Yasuda, K., Saito, Y., Nagata, K., and Hatayama, T. (2011). Characterization of stress sensitivity and chaperone activity of Hsp105 in mammalian cells. Biochem. Biophys. Res. Commun. *409*, 90–95.

Yao, J., Munson, K.M., Webb, W.W., and Lis, J.T. (2006). Dynamics of heat shock factor association with native gene loci in living cells. Nature *442*, 1050–1053.

Yoon, Y.J., Kim, J.A., Shin, K.D., Shin, D.-S., Han, Y.M., Lee, Y.J., Lee, J.S., Kwon, B.-M., and Han, D.C. (2011). KRIBB11 inhibits HSP70 synthesis through inhibition of heat shock factor 1 function by impairing the recruitment of positive transcription elongation factor b to the hsp70 promoter. Journal of Biological Chemistry *286*, 1737–1747.

Young, J.C., Hoogenraad, N.J., and Hartl, F.U. (2003). Molecular chaperones Hsp90 and Hsp70 deliver preproteins to the mitochondrial import receptor Tom70. Cell *112*, 41–50.

Zaret, K.S., and Carroll, J.S. (2011). Pioneer transcription factors: establishing competence for gene expression. Genes & Development *25*, 2227–2241.

Zheng, X., Krakowiak, J., Patel, N., Beyzavi, A., Ezike, J., Khalil, A.S., and Pincus, D. (2016). Dynamic control of Hsf1 during heat shock by a chaperone switch and phosphorylation. Elife *5*, 1153.

Zhong, M., Orosz, A., and Wu, C. (1998). Direct Sensing of Heat and Oxidation by Drosophila Heat Shock Transcription Factor. Molecular Cell *2*,

101–108.

Zhou, J., Wan, J., Gao, X., Zhang, X., Jaffrey, S.R., and Qian, S.-B. (2015). Dynamic m(6)A mRNA methylation directs translational control of heat shock response. Nature *526*, 591–594.

Zobeck, K.L., Buckley, M.S., Zipfel, W.R., and Lis, J.T. (2010). Recruitment timing and dynamics of transcription factors at the Hsp70 loci in living cells. Molecular Cell *40*, 965–975.

Zou, J., Guo, Y., Guettouche, T., Smith, D.F., and Voellmy, R. (1998). Repression of heat shock transcription factor HSF1 activation by HSP90 (HSP90 complex) that forms a stress-sensitive complex with HSF1. Cell *94*, 471–480.

Zuo, D., Subjeck, J., and Wang, X.-Y. (2016). Unfolding the Role of Large Heat Shock Proteins: New Insights and Therapeutic Implications. Frontiers in Immunology *7*, 267.

Zuo, J., Baler, R., Dahl, G., and Voellmy, R. (1994). Activation of the DNA-binding ability of human heat shock transcription factor 1 may involve the transition from an intramolecular to an intermolecular triple-stranded coiled-coil structure. Mol. Cell. Biol. *14*, 7557–7568.

CHAPTER 2[a]

MAMMALIAN HEAT SHOCK RESPONSE AND MECHANISMS UNDERLYING ITS GENOME-WIDE TRANSCRIPTIONAL REGULATION

Summary

The heat shock response (HSR) is critical for the survival of all organisms. However, it's scope, extent, and the molecular mechanism of regulation is poorly understood. Here we show that the genome-wide transcriptional response to heat-shock in mammals is rapid, dynamic, and results in induction of several hundred and repression of several thousand genes. Heat shock factor 1 (HSF1), 'the master regulator' of the HSR, controls only a fraction of the heat-shock induced genes, and does so by increasing RNA polymerase II (Pol II) release from promoter-proximal pause. The pervasive repression of transcription is predominantly HSF1-independent, and is mediated through reduction of Pol II pause-release. Moreover, the up- and down-regulated genes during HSR exhibit concomitant increase and decrease respectively in occupancy of pause-release factor positive transcription elongation factor b (P-TEFb). Interestingly, serum response factor (SRF) is

[a] Partially adapted from Mahat, D. B., Salamanca, H. H., Duarte, F. M., Danko, C. G., & Lis, J. T. (2016). Mammalian Heat Shock Response and Mechanisms Underlying Its Genome-wide Transcriptional Regulation. *Molecular Cell*, *62*(1), 63–78. Reprinted with permission from Elsevier.

transiently activated during the early phase of HSR, binds promoters of several cytoskeletal genes, and transiently induces their transcription in SRF-dependent manner. Overall, mammalian cells orchestrate rapid, dynamic, and extensive changes in transcription upon heat-shock that are largely modulated at pause-release, and HSF1 plays a limited and specialized role.

Introduction

Organisms have to cope with various kinds of stresses for survival, and one of the major repercussions of stress is the disruption of protein homeostasis. Without a proper rescue mechanism, accumulation of unfolded proteins and failure to properly fold newly synthesized polypeptides result in protein aggregates that interfere with cellular functions and ultimately leads to apoptosis (Milleron and Bratton, 2006). To protect cells from proteotoxic environment caused by heat stress, organisms deploy an evolutionarily conserved mechanism known as heat shock response (HSR). Exposure to elevated temperature, as well as many other stresses, triggers HSR, which is characterized by rapid and robust induction of heat shock protein genes (*Hsps*) (Berendes, 1968). HSPs are molecular chaperones responsible for maintaining protein homeostasis and are critical for survival during stress (Lindquist and Craig, 1988).

The HSR is orchestrated at the level of transcription by heat shock transcription factor (HSF) (Parker and Topol, 1984; Wu, 1984). Vertebrates

have four *HSF* genes, *Hsf1-4*. HSF1 is considered the master regulator of the HSR and is the ortholog of the sole *HSF* gene in invertebrates. HSF2 is the ubiquitously expressed HSF1 paralog that interplays with HSF1 during HSR and is involved in developmental pathways (Akerfelt et al., 2010; Rallu et al., 1997; Sarge et al., 1991). HSF3 and HSF4 show tissue-restricted expression and their roles in HSR remain to be explored (Akerfelt et al., 2010).

The current understanding of the mechanisms of transcription regulation during heat shock (HS) comes largely from in vitro and in vivo studies of the *Hsp70* gene in invertebrates (Morimoto, 1998). HSF1 is constitutively expressed as an inactive monomer and upon HS, HSF1 becomes active, trimerizes, (Westwood et al., 1991) and binds to the inverted repeats of nGAAn pentamers known as heat shock element (HSE) in the promoter of the *Hsp70* (Perisic et al., 1989; Sorger et al., 1987; Westwood et al., 1991). Thereupon, HSF1 recruits co-factors that dramatically increases the transcription of *Hsp70* (Anckar and Sistonen, 2011).

In addition to the induction of *Hsps*, HSF1 is implicated in various aspects of human physiology (Anckar and Sistonen, 2011). HSF1 plays an important role in aging and longevity (Morley and Morimoto, 2004; Murshid et al., 2013), protects organisms from obesity by regulating energy expenditure (Ma et al., 2015), and reduces susceptibility to stress in elderly hearts (Locke and Tanguay, 1996). More importantly, cancer cells co-opt HSF1 to support malignancy (Dai et al., 2007; Mendillo et al., 2012; Salamanca et al., 2014),

making its reduction in level or activity a potentially better target for cancer therapy than the several small inhibitory molecules against HSPs that are in ongoing clinical trials (Whitesell and Lindquist, 2009). In contrast, enhanced chaperone expression through clinical activation of HSF1 can improve prognosis of age-related neurodegenerative disorders that are caused by protein aggregates (Neef et al., 2011; 2010). This dichotomy in the role of HSF1 in cancer and neurological diseases requires a deeper understanding of the precise molecular mechanism of HSF1 driven gene regulation before clinical application of HSF1-based therapeutic tools.

Understanding networks of genes and cellular processes that are regulated during HSR is equally important to decipher how healthy cells survive acute stress and maintain proteostasis. Earlier genome-wide studies using microarray and RNA-sequencing indicated that additional genes besides *Hsps* are regulated during the HSR (Brown et al., 2014; Inouye et al., 2003; Page et al., 2006; Trinklein et al., 2004). These assays measure the cumulative effects of both transcriptional and post-transcriptional regulation, as they probe stable mRNA and lack the temporal resolution to reveal the first-order regulatory mechanisms of transcription. Transcription regulation consists of several steps any one of which might regulate gene expression, including Pol II recruitment, assembly of the initiation complex, promoter-proximal pausing and release from the pause, and Pol II elongation rates (Fuda et al., 2009). After more than 50 years since the discovery of HSR (Ritossa, 1962),

the breadth of transcriptional regulation during HSR, the precise step(s) modulated, and the kinetics and dynamics of the regulation remain to be fully understood.

In this study, we examined the HSR at transcriptional level - a primary and major point of regulation. To identify genome-wide changes in transcription, we used precision nuclear run-on sequencing (PRO-seq) (Kwak et al., 2013) - an assay that maps transcriptionally engaged Pol II at a nucleotide resolution by nascent RNA sequencing. We paired PRO-seq measurements in mouse embryonic fibroblasts (MEFs) derived from HSF1 knockout (*Hsf1*$^{-/-}$) mouse and its wild type (WT) littermate (McMillan et al., 1998) with HSF1-bound chromatin immunoprecipitation sequencing (ChIP-seq) to identify the genome-wide targets of HSF1 and its role in HSR. A time course of PRO-seq and HSF1 ChIP-seq prior to HS and immediately after HS extending up to an hour allowed us to observe the primary and secondary effects of heat-stress in transcription and decipher the roles of HSF1. We find HSF1 to be critical for induction of *Hsps*, other chaperones, and over 200 additional genes; however, the transcriptional changes during HSR are remarkably extensive and the majority of these changes are HSF1-independent. We identify different regulated classes of genes based on kinetics and dynamics of transcriptional regulation. We also decipher the mechanistic step in transcription where HSF1 exerts its regulation, as well as the mechanism of global repression. Together, these comprehensive and

highly sensitive analyses indicate that HSR is much more elaborate than previously understood, and regulators in addition to HSF1 are mobilized.

Materials and methods

Cell lines

Immortalized MEFs generated from *Hsf1*$^{-/-}$ and its wild type littermate mice (McMillan et al., 1998) were a gift from Ivor Benjamin. *Hsf1&2*$^{-/-}$ MEFs were generated by Christians ES and Le Dréan Y by crossing *Hsf1*$^{-/-}$ mouse and *Hsf2*$^{-/-}$ mice and the cells were immortalized by Valerie Mezger (Lecomte et al., 2010; 2013; McMillan et al., 1998; 2002).

Cell culture, heat shock, and nuclei isolation

MEFs were grown in 150mm TC-treated and gamma irradiated cell culture dish in Dulbecco's Modified Eagle Medium supplemented with 10% heat inactivated fetal bovine serum (v/v) and 1% Penicillin Streptomycin (v/v) at 37^0C with 5% CO_2 and 90% humidity.

Instantaneous HS was performed using ~80% confluent cells by adding pre-heated (42^0C) conditioned media collected from identically growing cells and by incubating the cell plates at 42^0C for the desired time. Heated media was used to ensure instantaneous HS as it takes more than 2.5 minutes for 37^0C cell culture plates to reach 42^0C in 42^0C incubator. Also, use of conditioned media from another plate of same cells grown identically avoids

complications that could result from the use of fresh media that contains fresh serum and higher glucose concentration.

After the desired duration of HS, NHS cells grown at 37^0C and the heat shocked cells were harvested identically. Nuclei were isolated as described previously with minor modifications (Core et al., 2008). Cells were rinsed once with ice-cold PBS and incubated in cell membrane lysis buffer (10mM Tris-Cl pH 8.0, 300mM sucrose, 3mM $CaCl_2$, 2mM $MgAC_2$, 0.1% TritonX-100, 0.5mM DTT) for five minutes. After incubation, cells were dounced 25 times using a dounce homogenizer (wheaton # 357546, loose pestle with 114nm clearance) and nuclei were harvested by centrifugation at 1000g for 5 mins. The isolated nuclei were washed twice in lysis buffer and resuspended in storage buffer (10 mM Tris-HCl pH 8.0, 25% glycerol, 5 mM $MgAc_2$, 0.1 mM EDTA, 5 mM DTT).

Nuclear run-on and PRO-seq library preparation

Nuclear run-on experiments were carried out as described previously with some modifications (Kwak et al., 2013). Briefly, 10×10^6 nuclei in 100μl of storage buffer were mixed with 100μl of 2x nuclear run-on buffer (10 mM Tris-HCl pH 8.0, 5 mM $MgCl_2$, 1 mM DTT, 300 mM KCl, 1% Sarkosyl, 50 uM biotin-11-A/C/G/UTP, 0.2 units/μl RNase inhibitor) and incubated at 37^0C for three minutes. RNAs were isolated and base-hydrolyzed with 200 nM final concentration of NaOH to an average size between 100-150 nucleotides. Nascent RNAs were isolated with magnetic beads coated with streptavidin

followed by 3' adapter ligation. After another round of biotin-streptavidin affinity purification, the mRNA cap was removed and the 5' adapter was ligated. After the third biotin-streptavidin affinity purification, cDNA was generated by reverse transcription and PRO-seq libraries were prepared for sequencing using Illumina TruSeq small-RNA adaptors with nine cycles of PCR.

Mapping of PRO-seq sequencing reads

Adapters from PRO-seq reads were clipped using cutadapt (Marcel, 2011). Remaining reads were trimmed to a maximum length of either 30, 32, 34, or 36bp. We found that 36bp reads result in the highest fraction mapping to the genome, and thus, the clipped reads were trimmed to 36bp for all downstream analysis. The filtered reads (>15bp and up to 36bp) were first mapped to a single copy of ribosomal DNA (45,309 bp long, TPA Accession No. BK0000964) to remove the contribution of nascent RNAs from ribosomal genes. The remaining reads were mapped to the mouse genome mm10 (released on December 2011) using bowtie (Langmead et al., 2009) and were required to uniquely map to the genome allowing up to 2 mismatches.

Normalization of PRO-seq libraries and validation of normalization

To account for the possibility of global change in transcription upon HS, we pursued different strategies to normalize data sets that are not based on using total reads. We tested the possibility of using histone genes and

ribosomal genes, but we found that transcription of many of these genes changes during HS. Ultimately, we discovered that the use of PRO-seq reads in the 3'end of long genes, where the advancing or receding wave of new pause-escaped Pol II upon HS would have not reached, is a robust, sensitive, and reliable strategy for normalization. This strategy also normalizes for all possible variations in sample handling and PRO-seq library preparation.

We selected 288 genes that are longer than 400kb for normalization (Figure 2.2C). Genes were further required to have at least 10x higher transcription than background regions, which are long gene-desert regions in different chromosomes, to avoid the inclusion of very lowly expressed genes in normalization. We reasoned that Pol II wave would not reach beyond 360 kb by 60 minutes of HS even at an exceptionally high rate of transcription (6kb/min). Thus, PRO-seq reads from 370kb downstream from the TSS to 10kb upstream of PAS were used for normalization. This novel approach to normalization was validated in three independent ways:

1. After normalization of libraries at all time points, we compared the change in PRO-seq density in the 5' and 3' end of significantly changed genes at 12'HS that are longer than 96 kb. At this earlier time point, the changes in Pol II distribution does not progress as far as in the 60'HS time point, allowing us to examine more genes and also exclude genes that were used for normalization. The 5' end (1 kb downstream of TSS up to 25 kb = 24 kb) region of the genes showed expected change in PRO-seq density, and

the 3' end (72 kb downstream of TSS to 96 kb = 24 kb, and as before, the rationale for avoiding the first 72 kb region is to avoid the newly released Pol II wave that could be present in this region - 12 minutes * very-high elongation rate of 6 kb/min = 72 kb) showed no change suggesting the normalization approach did not create any bias (Figure 2.2E).

2. We further used 52 genes that were identified as unchanged upon HS in MEFs out of 437 genes that were tested using an independent microarray experiment (Trinklein et al., 2004). Expression level of these genes after normalization was unbiased across the HS time course.
3. Finally, we examined the transcription level of 339 housekeeping genes not regulated following HS (la Grange et al., 2005) and found their normalized Pol II densities were unbiased across the HS time course.

Together, these analyses confirm that the normalization process was not biased and can be reliably used in PRO-seq time-course experiments in the future.

Differential expression analysis

DESeq2 (Love et al., 2014) was used to identify genes with differential PRO-seq density between NHS and HS libraries. In order to only consider the regions that likely have change in PRO-seq density, we used the window from 1 kb downstream of TSS up to 24 kb downstream of TSS or 1 kb upstream of PAS, whichever was shorter. The beginning of this 24 kb region started from 1

kb downstream of TSS in order to avoid the promoter-proximal paused Pol II, which is mostly in the first 100 bp but could appear to be further downstream due to some improperly annotated TSS. The 24 kb region was selected to ensure that Pol II had enough time to spread across most of the changed region at the conservative elongation rate of 2kb/min (Danko et al., 2013; Jonkers et al., 2014) * 12min = 24kb. To identify significantly changed genes at 2.5'HS, PRO-seq density from 500 bp downstream of TSS up to 5 kb downstream of TSS or 1 kb upstream of PAS, whichever was shorter, was used. For a gene to be called significantly changed between NHS and HS conditions, we set the PRO-seq density fold change threshold at 1.25 and DESeq2 p-value threshold at 0.001. Because of the DESeq2 threshold, we were unable to assign the mode of regulation to some genes. For example, in Figure 2F, 13% of the genes (100 - 17 - 53 - 4 - 13 = 13) bound by HSF1 in promoter are unaccounted because these genes did not meet the statistical threshold set by the differential expression analysis package.

Removing potential false positives

By visual inspection of genes in the genome browser, we discovered some genes that could be falsely identified to have differential PRO-seq density between NHS and HS conditions (Figure 2.3A). These false positives result from either run-over transcription from upstream genes (for example, *Abca6* could be called by DESeq2 to have increase in PRO-seq density at

60'HS compared to NHS when, in fact, this was the result of run-over transcription from the upstream gene *Abca5*), or intronic enhancers and internal TSSs (*Tmem88b* is not expressed which is evident by the lack of PRO-seq signal at the TSS of the gene but has an internal TSS or intronic enhancer evident by divergent transcription and the PRO-seq reads in the anti-sense direction of this region is being counted as the gene body PRO-seq reads of *Tmem88b*). These false positive genes are on the range of 18% of all sampled genes. To filter these genes, we used discriminative Regulatory Element detection from GRO-seq (dREG), a machine learning method that uses support vector regression to identify active transcriptional regulatory elements (Danko et al., 2015). To be considered for downstream analysis, we required the genes that have differential PRO-seq density in their gene body according to DESeq2 to have a dREG peak in their TSS. This approach very effectively filtered genes that could be falsely identified as being differentially regulated upon HS.

Western immunoblotting

Serial dilution of whole cell protein extract (20, 10, and 5 µg) from WT and *Hsf1*$^{-/-}$ MEFs was subjected to SDS-PAGE and transferred to nitrocellulose membrane. After blocking the membrane for 1 hour in 5% (w/v) bovine serum albumin (BSA), the membrane was probed with rabbit anti-HSF1 antibody (Cell Signaling #4356) and mouse anti-actin (Millipore #MAB1501)

overnight according to the manufacturer's instructions. Incubation with primary antibodies was followed by incubation in corresponding secondary antibodies (anti-rabbit conjugated to IRDye 800CW and anti-mouse conjugated to IRDye 680RD) for one hour and imaged in LI-COR Odyssey® CLx imaging system.

RNA-seq data and analysis

RNA-seq data in mouse 3T3 cells before and after HS that is used in Figures 2.3C and 2.3D was downloaded from NCBI Sequence Read Archive (accession number SRP035393). Sequences were aligned to the mouse genome using Tophat2 (Kim et al., 2013) and the mapped reads in each gene were quantified using HTSeq (Anders et al., 2015).

HSF1 and SRF ChIP-seq library preparation

ChIP was performed as described previously with some modifications (Guertin and Lis, 2010). 2.4 x 10^7 cells for each IP were cross-linked using 2% paraformaldehyde for five minutes at room temperature. The cross-linking reaction was quenched with 250 mM Glycine and the cell membrane lysed using Farnham lysis buffer (5mM PIPES pH 8.0, 85mM KCl, 0.5% NP-40, 1 tablet/50ml Protease Inhibitor). Nuclei were pelleted by centrifugation and incubated in RIPA buffer (1x PBS, 1% NP-40, 0.5% Sodium Deoxycholate, 0.1% SDS, 1 tablet/50ml Protease Inhibitor) for 20 minutes and sonicated for 45 minutes in the highest setting in a Bioruptor® (15s ON 30 sec OFF).

Primary antibodies (first HSF1 antibody (Ab1) from Cell Signaling #4356, second HSF1 antibody (Ab2) gift from Richard Morimoto Lab, and SRF antibody from Cell Signaling #5147) conjugated to magnetic beads were used to immunoprecipitate chromatin fragments and washed five times with Lithium Chloride (LiCl) wash buffer (100mM Tris pH 7.5, 500mM LiCl, 1% NP-40, 1% DOC). Samples were incubated overnight at 65^0C using elution buffer (1% SDS and 100mM $NaHCO_3$) to reverse crosslinks.

The ends of purified DNA were repaired and dATP was ligated to the 3' ends. Illumina TRUseq DNA adapters with dTTP overhang were ligated on both ends of the DNA and ChIP-seq libraries, and were PCR amplified for five initial cycles. The libraries were gel purified to remove adapter dimers and further PCR amplified for an additional four cycles.

ChIP-seq peak calling and combining p-values of overlapping HSF1 peaks from two different antibodies

HSF1 peaks in ChIP-seq libraries prepared with two HSF1 antibodies were called against library prepared with non-specific IgG using MACS (Zhang et al., 2008). In order to minimize the false positive peaks, we only considered HSF1 peaks that were identified with both antibodies at a threshold of $1e^{-03}$. We then combined the two p-values of a peak from two antibodies using Fisher method for combining two p-values for effects in the same direction ([Fisher, 1958] Fisher, R. A. (1958). Statistical Methods for Research Workers,

13th Edition. Hafner Publishing). We selected HSF1 peaks that passed a combined p-value threshold of $1.0e^{-08}$, which is equivalent to two individual p-values of $1e^{-05}$ (default p-value used by MACS). We further removed HSF1 peaks that were present in *Hsf1*$^{-/-}$ MEFs as we assumed that peaks identified in *Hsf1*$^{-/-}$ ChIP-seq libraries were false positives.

HSE motif

Empirically identified position weight matrix of HSE from HSF1 ChIP-seq peaks was scanned over the entire mouse genome using FIMO from the MEME suite (Grant et al., 2011). Motifs that have p-values less than $1e^{-05}$ were used to score the presence of HSE in the genome.

GO analysis

Molecular functions and biological processes enriched in HS regulated genes were identified using DAVID (the database for annotation, visualization and integrated discovery) (Huang et al., 2009). Highly enriched and non-redundant GO terms from the PANTHER database were selected using the option provided in DAVID.

Enrichment of TF binding motifs

Position weight matrices corresponding to known TFs curated from several repositories (Jolma et al., 2013; Mathelier et al., 2014; Neph et al.,

2012; Trinklein et al., 2004; Weirauch et al., 2014) were first clustered based on DNA sequence recognition site similarity and scanned in the promoters of regulated genes using the RTFBSDB package (Kwak et al., 2013). The p-value was calculated against the 3rd order Markov model background constructed from all active promoters in MEFs.

ENCODE and non-ENCODE genomic data

ChIP-seq data for histone modifications and Pol II ser5P were downloaded from GEO. H3K4me1, H3K4me3, and H3k27ac were generated by ENCODE (ENCODE Project Consortium, 2012), H3K9me3 and H3K79me2 were generated by the Richard Young lab and Pol II ser5P by Lee JT lab.

Measurement of elongating Pol II wave

The Pol II wave was identified in upregulated genes at 2.5'HS and 12'HS, and also in downregulated genes at 12'HS and 60'HS. A three-state hidden Markov model, with states representing the 5' upstream region, the Pol II wave, and the 3' end of the gene, were used to identify the start and end of the PRO-seq waves (Danko et al., 2013). Differences in PRO-seq read counts in 50bp windows between NHS and HS time points in upregulated genes at 2.5'HS and 12'HS were used as input for HMM (Figures 3E-H), whereas 2 kb windows were used for downregulated genes at 12'HS and 60'HS (Figure S4G).

Results

HS triggers rapid, robust, and diverse changes in transcription

To characterize the global changes in transcription associated with HSR and to understand the role of HSF1, we performed genome-wide PRO-seq assays on WT and *Hsf1*$^{-/-}$ MEFs. We prepared two biological replicates of PRO-seq libraries at 37^{0}C (NHS) and 2.5, 12, and 60 minutes after an instantaneous HS at 42^{0}C (Figure 2.1A). The libraries were sequenced to high depth (Table 2.1) and mapped to the mouse genome (mm10). The biological replicates correlated well (Figure 2.1B and Table 2.2), and as expected, the *Hsf1*$^{-/-}$ MEFs produced no HSF1 protein (Figure 2.1C), nor any PRO-seq reads in the deleted region of the *Hsf1* gene (Figure 2.1D).

Normalization of genomic libraries by conventional methods such as total mapped reads or ribosomal RNA reads are inadequate when dealing with significant changes in total transcription. Therefore, we devised a novel approach for normalizing the PRO-seq libraries using PRO-seq reads from the 3' end of very long genes (>400 kb), the regions beyond the advancing or receding wave of Pol II even at the longest HS time point (60'HS) (Figures 2.1E and 2.1F). This normalization approach was validated using three different tests. First, the PRO-seq density after normalization in the 3' ends of significantly upregulated and downregulated genes after a 12'HS is unchanged, while the 5' ends show expected change (Figures 2.2A – note this includes a larger collection of genes than used for the normalization and the

Figure 2.1. A novel and reliable normalization approach for PRO-seq
(A) Experimental set-up, PRO-seq assay was performed in nuclei isolated from WT and *Hsf1*$^{-/-}$ MEFs at NHS, 2.5'HS, 12'HS, and 60'HS.
(B) A representative spearman correlation plot between the biological replicates of PRO-seq libraries in the pause region (left) and the gene-body region. Color of hex-bins represents the number of reads in each bin. Complete table of Pearson and Spearman correlation values are listed in Table 2.2.
(C) Quantitative Western Blot analysis of HSF1 in WT and *Hsf1*$^{-/-}$ MEFs. Actin is used as the loading control.
(D) Genome browser screenshot of *Hsf1* gene. The 2 kb region devoid of PRO-seq reads (green box) represents half of the DNA binding domain and trimerization domain excised to create *Hsf1*$^{-/-}$ MEFs.
(E) Graphical representation of the genes used for normalization of PRO-seq libraries. A region that is at least 20 kb covering an interval from 370 kb downstream of TSS to 10 kb upstream from PAS in the 3' end of genes longer than 400 kb is used for normalization.
(F) Screenshot of two long genes used for normalization. *Zfpm2* is upregulated and *Exoc6b* is downregulated at 60'HS, which is evident by increase and decrease in PRO-seq density in the 5'end of genes respectively. The green box represents the 3' end region of the genes used for normalization.

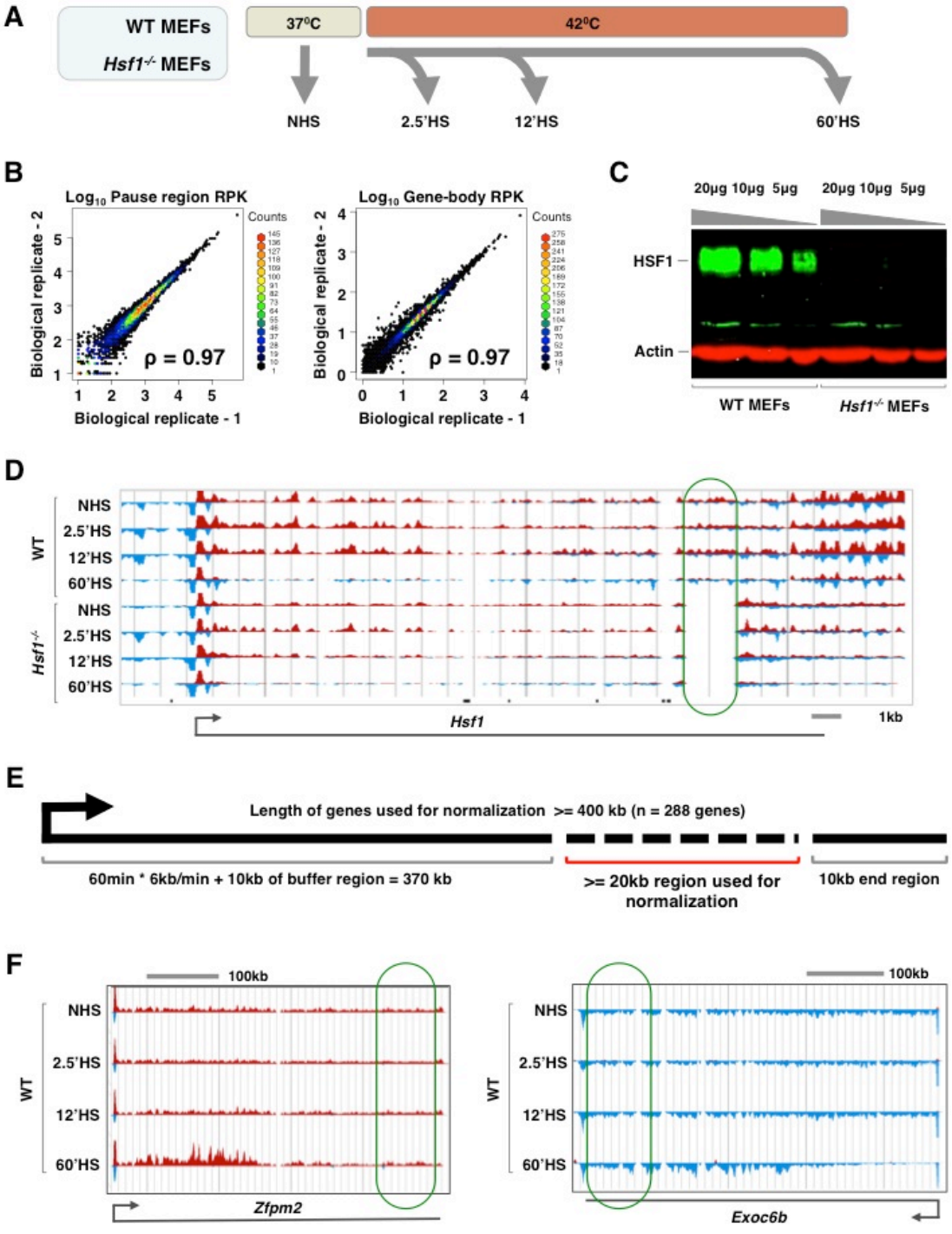

A
WT MEFs
Hsf1-/- MEFs
37°C
42°C
NHS
2.5'HS
12'HS
60'HS
B
Log10 Pause region RPK
Log10 Gene-body RPK
Counts
Biological replicate - 2
Biological replicate - 1
ρ = 0.97
C
20μg 10μg 5μg
20μg 10μg 5μg
HSF1
Actin
WT MEFs
Hsf1-/- MEFs
D
WT
Hsf1-/-
NHS
2.5'HS
12'HS
60'HS
Hsf1
1kb
E
Length of genes used for normalization >= 400 kb (n = 288 genes)
60min * 6kb/min + 10kb of buffer region = 370 kb
>= 20kb region used for normalization
10kb end region
F
100kb
WT
NHS
2.5'HS
12'HS
60'HS
Zfpm2
Exoc6b

Figure 2.2. Validation of normalization and use of dREG to minimize potential false positives

(A) PRO-seq density in the 5' end (region 1 kb downstream of TSS up to 25 kb) and the 3' end (24 kb region after the first 72 kb) of the upregulated genes (left) and downregulated genes (right) at 12'HS. Genes used in this analysis does not include genes used for normalization and are longer than 96 kb.

(B) Correlation of normalized PRO-seq density before and after HS in genes not bound by HSF1 as measured by ChIP-chip and not induced upon HS as measured by microarray (n=52) after HS (Trinklein et al., 2004). The red line represents the linear fit of all data points. A complete list of spearman correlation values for these genes between NHS and HS in WT and *Hsf1*$^{-/-}$ MEFs are listed in Table 2.3.

(C) Correlation of normalized PRO-seq density before and after HS in house-keeping genes (n=339). The red line represents the linear fit of all data points. A complete list of spearman correlation values for these genes between NHS and HS in WT and *Hsf1*$^{-/-}$ MEFs are listed in Table 2.3.

(D) Fraction of genes that are true positive and have clean transcription units with dREG peak overlapping the TSS region of the genes, run-over transcription from the upstream genes, and the genes with internal TSS or intronic enhancer. For instance, *Abca6* is a result of run-over transcription from upstream *Abca5* gene in the same strand. This gene would be called significantly upregulated in 60'HS compared to NHS. Similarly, *Tmem88b* has an internal TSS or intronic enhancer that contributes to gene-body PRO-seq density and would also be called significantly downregulated at 12'HS compared to NHS.

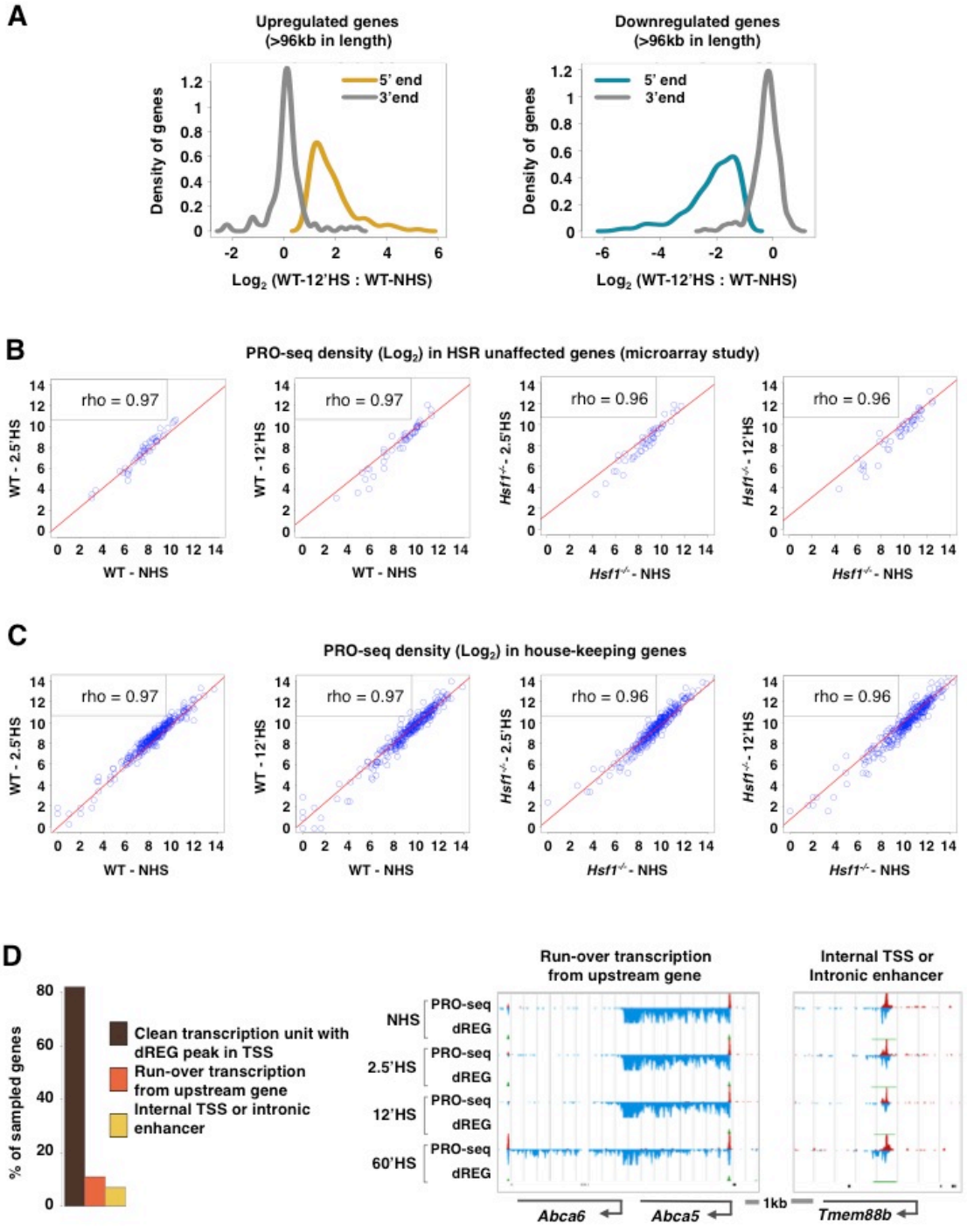

A
Upregulated genes
(>96kb in length)
Density of genes
5' end
3'end
Log2 (WT-12'HS : WT-NHS)
Downregulated genes
(>96kb in length)
Density of genes
5' end
3'end
Log2 (WT-12'HS : WT-NHS)
B
PRO-seq density (Log2) in HSR unaffected genes (microarray study)
rho = 0.97
WT - 2.5'HS
WT - NHS
rho = 0.97
WT - 12'HS
WT - NHS
rho = 0.96
Hsf1-/- - 2.5'HS
Hsf1-/- - NHS
rho = 0.96
Hsf1-/- - 12'HS
Hsf1-/- - NHS
C
PRO-seq density (Log2) in house-keeping genes
rho = 0.97
WT - 2.5'HS
WT - NHS
rho = 0.97
WT - 12'HS
WT - NHS
rho = 0.96
Hsf1-/- - 2.5'HS
Hsf1-/- - NHS
rho = 0.96
Hsf1-/- - 12'HS
Hsf1-/- - NHS
D
% of sampled genes
Clean transcription unit with dREG peak in TSS
Run-over transcription from upstream gene
Internal TSS or intronic enhancer
Run-over transcription from upstream gene
Internal TSS or Intronic enhancer
NHS
2.5'HS
12'HS
60'HS
PRO-seq
dREG
Abca6
Abca5
1kb
Tmem88b

Table 2.1. Sequencing depth and alignment statistics of PRO-seq in WT and *Hsf1*$^{-/-}$ MEFs. Significant fraction of reads were not uniquely mapped to the genome because of the deliberate inclusion of short fragments in the final library for sequencing in order to optimize the detection of promoter-proximal paused Pol II. Libraries were sequenced in Illumina HiSeq-2000.

Library	Sequenced	Ribo-somal (%)	Un-aligned (%)	Non-unique (%)	Aligned (mm10)
WT_NHS_BR1	116,597,845	11.80	21.48	16.22	21,942,702
WT_NHS_BR2	74,543,714	9.92	36.19	13.81	14,069,970
WT_2.5'HS_BR1	124,769,718	14.33	7.06	21.53	16,889,583
WT_2.5'HS_BR2	113,241,306	14.26	6.74	20.79	13,776,348
WT_12'HS_BR1	154,121,332	19.25	8.69	19.26	23,254,910
WT_12'HS_BR2	87,985,404	19.57	16.72	24.32	17,678,637
WT_60'HS_BR1	153,916,793	6.61	20.11	23.39	30,211,715
WT_60'HS_BR2	17,742,544	6.77	37.62	23.79	4,222,282
***Hsf1*$^{-/-}$_NHS_BR1**	71,252,588	18.66	11.52	14.43	33,556,982
***Hsf1*$^{-/-}$_NHS_BR2**	65,940,822	20.25	11.05	15.81	30,059,980
***Hsf1*$^{-/-}$_2.5'HS_BR1**	37,020,268	13.57	4.16	26.76	6,442,517
***Hsf1*$^{-/-}$_2.5'HS_BR2**	39,429,009	10.66	3.69	47.85	5,979,212
***Hsf1*$^{-/-}$_12'HS_BR1**	55,063,106	20.03	7.44	20.26	18,035,093
***Hsf1*$^{-/-}$_12'HS_BR2**	90,945,600	17.87	7.30	20.48	26,820,803
***Hsf1*$^{-/-}$_60'HS_BR1**	28,318,086	9.13	5.74	23.60	11,497,974
***Hsf1*$^{-/-}$_60'HS_BR2**	60,676,032	9.57	7.01	23.24	23,054,099

Table 2.2. PRO-seq biological replicates are highly correlated. PRO-seq reads for each gene were correlated for promoter region (100bp region in the 5' end of the genes) and gene body region (500 bp downstream of TSS to polyA site).

	Pearson Correlation		Spearman correlation	
	Promoter	**Gene body**	**Promoter**	**Gene body**
WT_NHS	0.93	0.99	0.97	0.97
WT_2.5'HS	0.99	0.99	0.98	0.97
WT_12'HS	0.96	0.99	0.97	0.97
WT_60'HS	0.94	0.99	0.95	0.94
***Hsf1*$^{-/-}$_NHS**	0.99	0.99	0.98	0.97
***Hsf1*$^{-/-}$_2.5'HS**	0.91	0.99	0.97	0.97
***Hsf1*$^{-/-}$_12'HS**	0.99	0.99	0.98	0.97
***Hsf1*$^{-/-}$_60'HS**	0.99	0.99	0.97	0.96

Table 2.3. Normalization of PRO-seq libraries using PRO-seq reads in the 3'end of long genes show expected results. Spearman correlation for previously identified genes as not bound by HSF1 and not changed upon HS in microarray study from Richard Myers lab (Trinklein et al., 2004) (top), and housekeeping genes from previously published study (la Grange et al., 2005) (bottom) show absence of systematic bias in normalization scheme used in this study. We would expect to see a systematic bias if the normalization method was non-optimal.

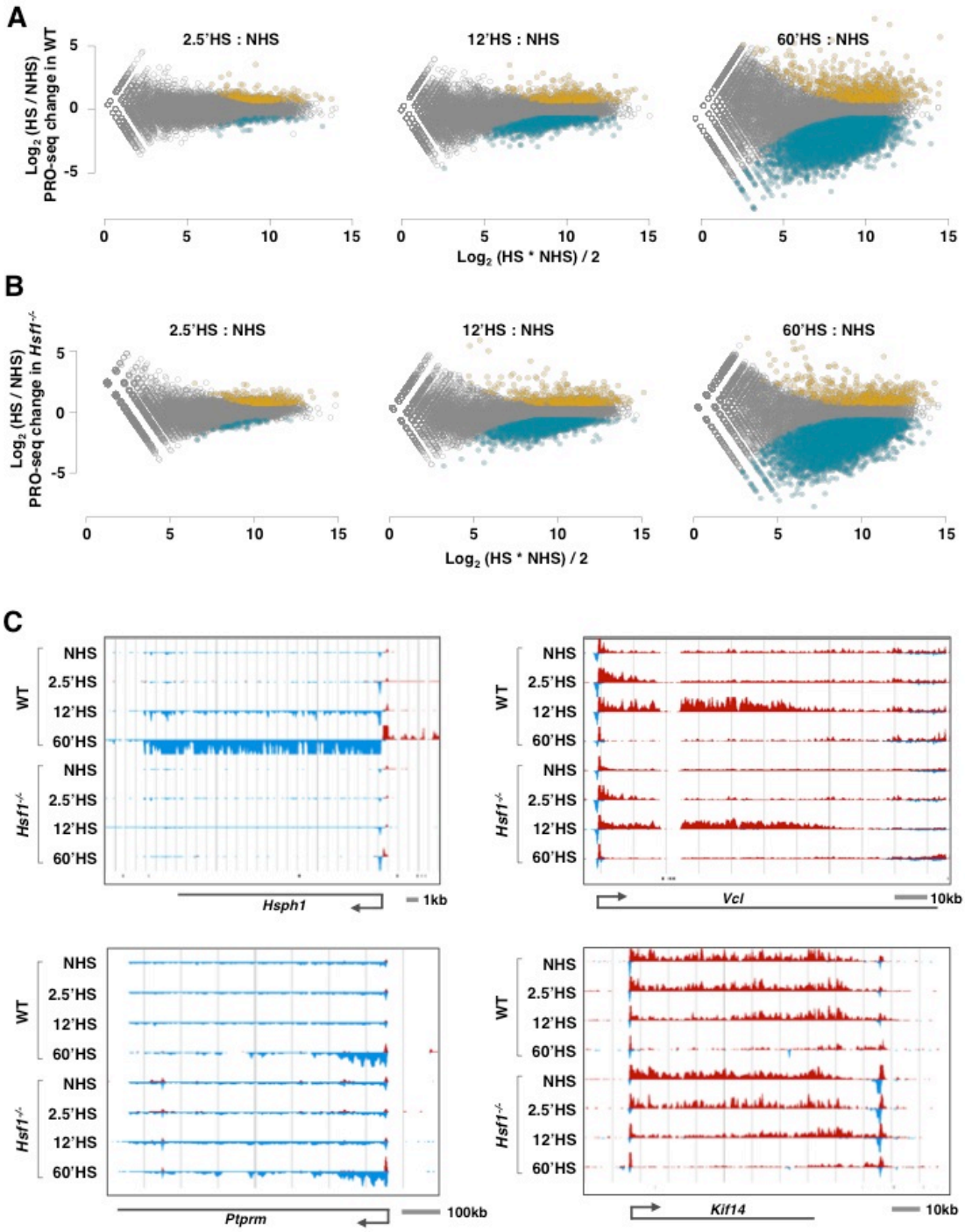
	2.5'HS vs NHS	12'HS vs NHS	60'HS vs NHS
WT	0.97	0.97	0.72
***Hsf1*$^{-/-}$**	0.96	0.96	0.84

	2.5'HS vs NHS	12'HS vs NHS	60'HS vs NHS
WT	0.97	0.97	0.81
***Hsf1*$^{-/-}$**	0.96	0.96	0.84

genes used for normalization are excluded in this test). Second, a set of genes identified as unaffected during HSR in MEFs using microarrays (Trinklein et al., 2004) showed no changes in PRO-seq density between NHS and HS conditions after normalization (Figure 2.2B & Table 2.3). Third, a previously defined group of house-keeping genes (la Grange et al., 2005) also showed no systematic deviation between HS and NHS conditions (Figure 2.2C & Table 2.3). After normalization, genes that could be falsely detected in differential expression analysis due to a) transcription running past the 3' end of upstream genes and b) internal TSS or intronic enhancers were eliminated using dREG (Danko et al., 2015) (Figure 2.2D) (see methods).

To characterize the transcriptome regulated during HS, we used DESeq2 (Love et al., 2014) to identify differential PRO-seq density in the body of genes before and after HS. A substantial fraction of the transcriptome changes upon HS, and the number of genes and the level of change progressively increase with time (Figures 2.3A and 2.3B). Moreover, the kinetics and dynamics of change is remarkably diverse (Figure 2.3C). First, many *Hsps* are robustly and persistently induced in an HSF1-dependent manner (such as *Hsph1* with ~ 60 fold induction). Second, many genes are immediately and transiently induced (like *Vcl*) where the advancing wave of newly transcribing Pol II is particularly noticeable, and this induction is independent of HSF1. Third, many genes show late induction (such as *Ptprm*), some of which could be targets of late-activated HSF1 or transcription factors

Figure 2.3. HS induces rapid, dynamic, and extensive changes in transcription in a mostly HSF1-independent manner
(A) 'Minus-average' (MA) plots represent PRO-seq density change in the gene-body of all genes (n=23460) between NHS and 2.5'HS (left panel), 12'HS (mid panel), and 60'HS (right panel) in WT MEFs. Significantly upregulated genes (p-value < 0.001 in DESeq2) are shown in gold and significantly downregulated genes are shown in blue.
(B) MA plots represent PRO-seq density change in the gene-body of all genes (n=23460) between NHS and 2.5'HS (left panel), 12'HS (mid panel), and 60'HS (right panel) in *Hsf1*$^{-/-}$ MEFs. Significantly upregulated genes are shown in gold and significantly downregulated genes are shown in blue.
(C) Screenshots of four genes with different kinetics and dynamics of regulation. PRO-seq density in sense and antisense direction is shown in red and blue respectively. Small vertical black bars at the bottom of PRO-seq tracks represent the genomic regions that do not map uniquely at 36bp resolution.

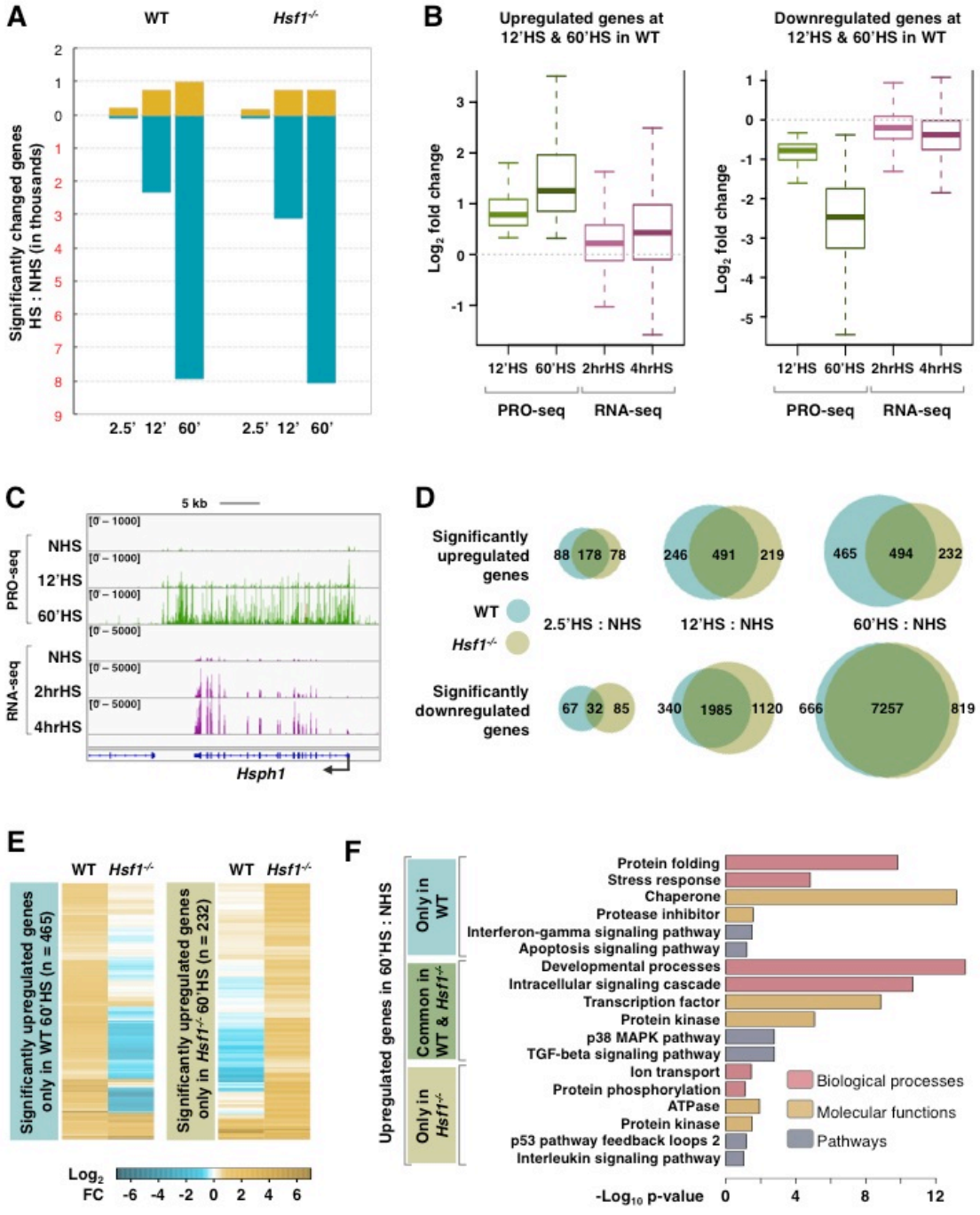

A
2.5'HS : NHS
12'HS : NHS
60'HS : NHS
Log2 (HS / NHS)
PRO-seq change in WT
Log2 (HS * NHS) / 2
B
2.5'HS : NHS
12'HS : NHS
60'HS : NHS
Log2 (HS / NHS)
PRO-seq change in Hsf1-/-
Log2 (HS * NHS) / 2
C
WT
Hsf1-/-
NHS
2.5'HS
12'HS
60'HS
Hsph1
1kb
Vcl
10kb
Ptprm
100kb
Kif14
10kb

(TFs) induced early during HS, and the majority of these are independent of HSF1. Fourth, a large fraction of the expressed genes (like *Kif14*) are significantly downregulated, and all or nearly all are independent of HSF1. Overall, DESeq2 identifies significant upregulation of 10% and downregulation of 55% of all active genes (Figure 2.4A). For the majority of these genes, the change in transcription measured by PRO-seq is recapitulated at the mRNA level measured by RNA-seq (Shalgi et al., 2014), despite the fundamental difference between the two assays and the additional regulation of mRNA stability being a part of RNA-seq measurement (Figure 2.4B). Fold change in RNA-seq requires higher change in transcription than required for PRO-seq due to the higher level of steady-state mRNA level compared to the nascent transcription in a cell, which is reflected here by the lower magnitude of change in RNA-seq than in PRO-seq for the HS regulated genes (Figures 2.4B and 2.4C). Thus, our results indicate that the RNA regulation in response to HS is manifested directly at the level of transcription and consists of multiple distinct regulatory programs that are captured here with high spatiotemporal resolution and sensitivity afforded by the PRO-seq assay.

Majority of the HS-regulated genes are HSF1-independent

In addition to the unexpectedly higher number of changed genes, the similar numbers of regulated genes in WT and *Hsf1*$^{-/-}$ MEFs (Figure 2.4A) and the extent of their overlap are uncanny (Figures 2.4D). The overlap reported

Figure 2.4. Majority of the HS-regulated genes are HSF1-independent
(A) Number of significantly changed genes upon HS. Upregulated genes are shown in gold and downregulated genes are shown in blue.
(B) Comparison of PRO-seq density fold change and RNA-seq fold change in genes that are significantly upregulated (left) and downregulated (right) at 12'HS and 60'HS by PRO-seq measurement.
(C) Screenshot of *Hsph1* gene showing the PRO-seq and RNA-seq reads. The scale on y-axis is shown in the top left corner.
(D) Venn diagram of significantly upregulated and downregulated genes between WT and *Hsf1*$^{-/-}$ MEFs.
(E) PRO-seq density fold change in WT and *Hsf1*$^{-/-}$ MEFs at 60'HS in genes that are called significantly upregulated only in WT 60'HS (left) and only in *Hsf1*$^{-/-}$ 60'HS (right) by DESeq2.
(F) GO analysis of the three classes of genes upregulated at 60'HS. The GO terms in each gene class are further categorized into three groups of biological processes, molecular functions, and signaling pathways. The length of the bar (x-axis) denotes the p-value.


A
WT
Hsf1-/-
Significantly changed genes HS : NHS (in thousands)
2.5' 12' 60' 2.5' 12' 60'
B
Upregulated genes at 12'HS & 60'HS in WT
Downregulated genes at 12'HS & 60'HS in WT
Log2 fold change
12'HS 60'HS 2hrHS 4hrHS
PRO-seq RNA-seq
C
5 kb
PRO-seq
NHS
12'HS
60'HS
RNA-seq
NHS
2hrHS
4hrHS
Hsph1
D
Significantly upregulated genes
WT
Hsf1-/-
88 178 78
246 491 219
465 494 232
2.5'HS : NHS
12'HS : NHS
60'HS : NHS
Significantly downregulated genes
67 32 85
340 1985 1120
666 7257 819
E
WT Hsf1-/-
Significantly upregulated genes only in WT 60'HS (n = 465)
Significantly upregulated genes only in Hsf1-/- 60'HS (n = 232)
Log2 FC -6 -4 -2 0 2 4 6
F
Upregulated genes in 60'HS : NHS
Only in WT
Common in WT & Hsf1-/-
Only in Hsf1-/-
Protein folding
Stress response
Chaperone
Protease inhibitor
Interferon-gamma signaling pathway
Apoptosis signaling pathway
Developmental processes
Intracellular signaling cascade
Transcription factor
Protein kinase
p38 MAPK pathway
TGF-beta signaling pathway
Ion transport
Protein phosphorylation
ATPase
Protein kinase
p53 pathway feedback loops 2
Interleukin signaling pathway
Biological processes
Molecular functions
Pathways
-Log10 p-value 0 4 8 12

here in upregulated genes is likely an underestimation as many of the uniquely upregulated genes in WT MEFs are upregulated in *Hsf1*$^{-/-}$ MEFs and vice versa, but don't meet the DESeq2 threshold. For example, 21% of the 465 genes identified by DESeq2 as upregulated only in WT 60'HS also show upregulation in *Hsf1*$^{-/-}$ (Figure 2.4E), and similarly, 22% of the 232 genes identified by DESeq2 as upregulated only in *Hsf1*$^{-/-}$ 60'HS are also upregulated in WT, but they don't meet the statistical threshold of DESeq2. Compared to the upregulated genes, the downregulated genes show even more overlap between WT and *Hsf1*$^{-/-}$ MEFs indicating that transcriptional repression occurs by mechanisms that are largely HSF1-independent.

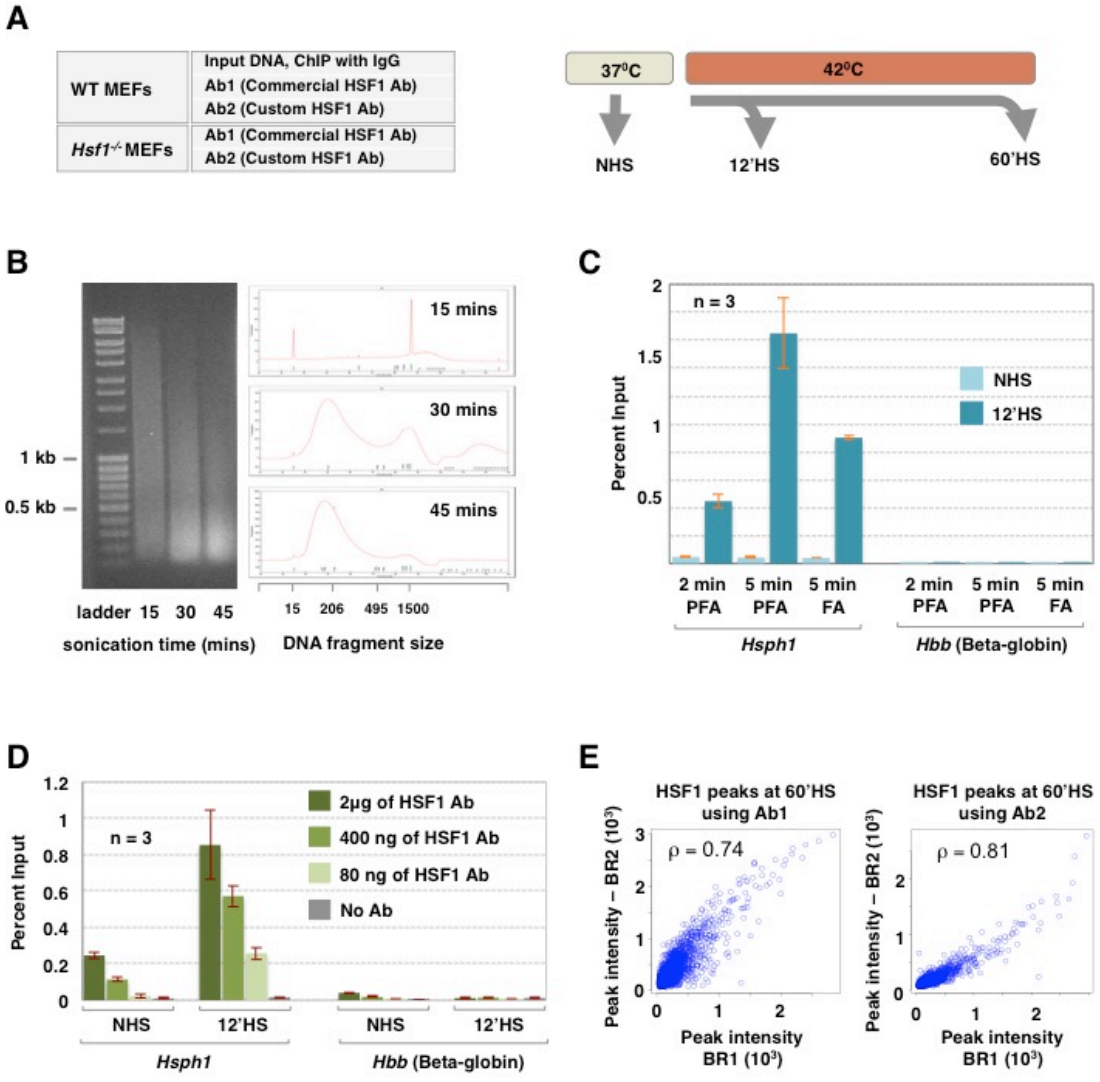
To examine the biological functions related to the genes that are differentially expressed in WT and *Hsf1*$^{-/-}$ MEFs, we performed gene ontology (GO) analysis using DAVID (Huang et al., 2009). As expected, genes upregulated only in WT MEFs are enriched for chaperones involved in protein folding and stress response (Figure 2.4F). Genes upregulated only in *Hsf1*$^{-/-}$ MEFs are enriched for ATPase and protein kinases. Genes that are upregulated in both cell types, which is a significant fraction of the upregulated genes, are enriched for transcription factors (TFs) and protein kinases involved in developmental processes. Overall, more than 87% of genes regulated at 60'HS in WT are similarly regulated in *Hsf1*$^{-/-}$ MEFs, and these observations suggest that alternative HSF1-independent mechanisms mediate widespread changes in transcription upon HS.

HSF1 binds to the promoters of a small fraction of HS-induced genes

Despite being called the master regulator of HSR, the majority of HS regulated genes appear to be HSF1-independent (Figures 2.4A & 2.4D). Therefore, to understand the role of HSF1 binding during HSR, we performed HSF1 ChIP-seq in WT and *Hsf1*$^{-/-}$ MEFs at NHS, 12'HS, and 60'HS (Figure 2.5A). We optimized various parameters like sonication (Figure 2.5B), crosslinkers and cross-linking duration (Figure 2.5C), and antibody concentrations (Figure 2.5D). To minimize false-positive peaks in ChIP-seq (Bailey et al., 2013; Chen et al., 2012; Pickrell et al., 2011), we used two different antibodies for ChIP that recognize different parts of HSF1 and prepared two biological replicates of ChIP-seq libraries for each antibody. As negative controls, we made ChIP-seq libraries from Input DNA, chromatin immuno-precipitated with non-specific antibody (IgG) in WT MEFs, and chromatin immuno-precipitated with both HSF1 antibodies in *Hsf1*$^{-/-}$ MEFs (Table 2.4). Biological replicates correlated well (Figure 2.5E) and were combined, and ChIP-seq peaks were called using MACS (Zhang et al., 2008).

As expected, we find prominent HSF1 peaks in the promoters of classical *Hsps,* and the two antibodies generated similar ChIP-seq profile (Figure 2.6A). HSF1 peaks identified here are highly specific, 89% of the HSF1 bound sites contain canonical HSE (p-value < 0.00001) (Figures 2.6B and 2.6C), and the fold enrichment of HSF1 correlates with the motif match score

Figure 2.5. Optimization of various parameters for HSF1 ChIP-seq
(A) Experimental set-up, ChIP-seq libraries in WT MEFs were made in duplicates with Input DNA, chromatin immunoprecipitated with non-specific IgG and two different HSF1 specific antibodies – Ab1 and Ab2 - at NHS, 12'HS, and 60'HS. Additionally, ChIP-seq libraries were also made with both HSF1 specific antibodies in *Hsf1*$^{-/-}$ MEFs.
(B) Distribution of DNA fragments size as a function of sonication time. Sonicated DNA was ran in agarose gel (left) and also assessed by automated sizing and quantitation using Agilent bioanalyzer kit (right).
(C) HSF1 ChIP-qPCR at *Hsph1* (positive) and *Hbb* (negative) gene before and after HS in order to test the efficiency of two crosslinking reagents (paraformaldehyde - PFA and formaldehyde - FA) and crosslinking time. Y-axis represents percentage input and standard deviation is calculated from three independent experiments.
(D) HSF1 ChIP-qPCR at *Hsph1* (positive) and *Hbb* (negative) gene before and after HS in order to optimize different amounts of antibody for optimal result in ChIP-seq experiment. Y-axis represents percentage input and standard deviation is calculated from three experiments.
(E) Correlation of the intensity of overlapping HSF1 peaks between two biological replicates in WT 60'HS for Ab1 (left) and Ab2 (right).


A
WT MEFs
Input DNA, ChIP with IgG
Ab1 (Commercial HSF1 Ab)
Ab2 (Custom HSF1 Ab)
Hsf1-/- MEFs
Ab1 (Commercial HSF1 Ab)
Ab2 (Custom HSF1 Ab)
37°C
42°C
NHS
12'HS
60'HS
B
1 kb
0.5 kb
15 mins
30 mins
45 mins
ladder 15 30 45
sonication time (mins)
15 206 495 1500
DNA fragment size
C
n = 3
Percent Input
NHS
12'HS
2 min PFA
5 min PFA
5 min FA
Hsph1
Hbb (Beta-globin)
D
n = 3
Percent Input
2µg of HSF1 Ab
400 ng of HSF1 Ab
80 ng of HSF1 Ab
No Ab
NHS
12'HS
Hsph1
Hbb (Beta-globin)
E
HSF1 peaks at 60'HS using Ab1
ρ = 0.74
Peak intensity – BR2 (10³)
Peak intensity BR1 (10³)
HSF1 peaks at 60'HS using Ab2
ρ = 0.81

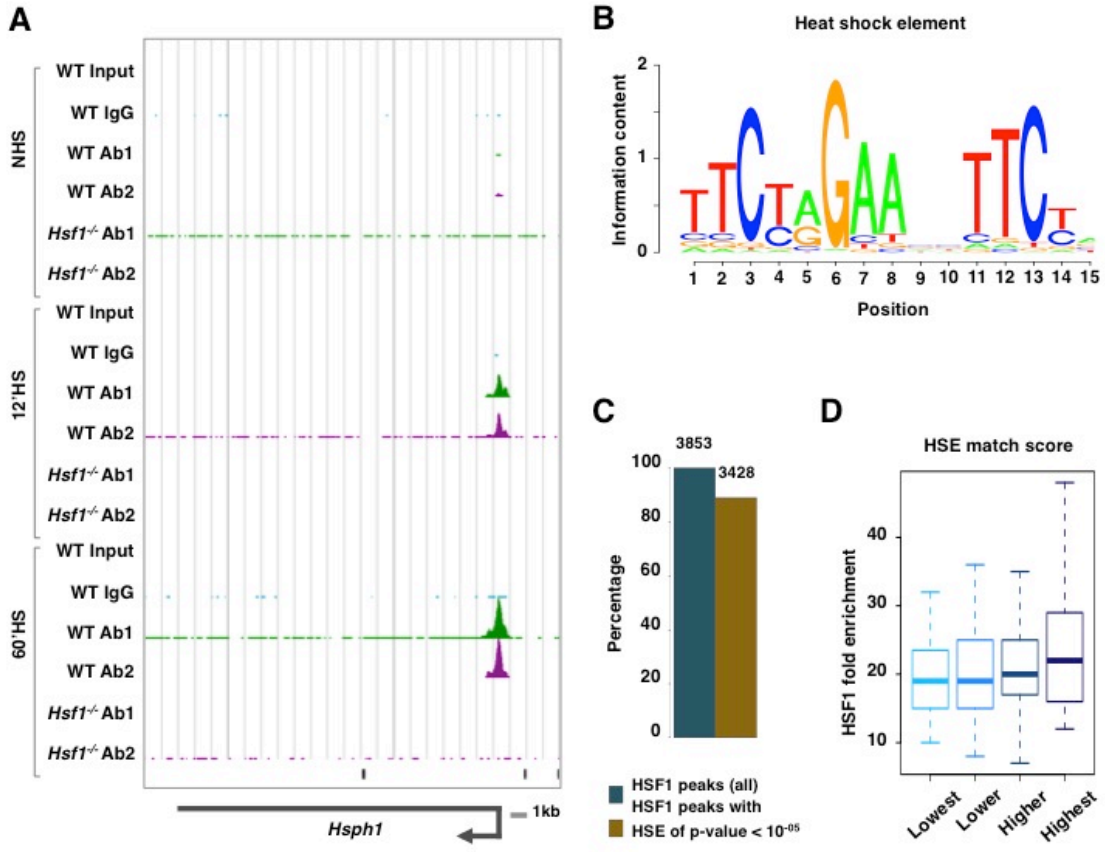


Figure 2.6. Use of two HSF1 antibodies yield high-quality ChIP-seq libraries

(A) Screenshot of *Hsph1* gene shows ChIP-seq read density in WT and *Hsf1*$^{-/-}$ MEFs. The scale on the y-axis represents the number of ChIP-seq tags under the peaks.

(B) Sequence logo of the most represented motif in HSF1 peaks identified by *de novo* motif search using MEME.

(C) Barplot showing the fraction of all HSF1 peaks that contains highly significant HSE ($p < 0.00001$) in the DNA sequence occupied by the peaks. Actual number of peaks in each case is shown on top.

(D) Fold enrichment of HSF1 peaks in the four quartiles classified by low to high match score to HSE (n=36, 36, 36, 35 respectively for the lowest, lower, higher, and highest).

Table 2.4. Sequencing depth and alignment statistics of HSF1 ChIP-seq libraries. ChIP-seq reads that pass the Illumina quality filter were retained, adapters were clipped and the reads longer than 15 bp were mapped to the mouse genome.

Library	Sequenced	Adaptor dimers	Unaligned	Uniquely aligned
WT_NHS_Input	8,015,869	997,080	309,635	6,709,154
WT_NHS_IgG_BR1	8,206,235	2,676,621	232,390	5,297,224
WT_NHS_IgG_BR2	9,721,219	3,562,286	654,125	5,504,808
WT_NHS_HSF1_CS_BR1	28,188,028	5,766,142	1,006,461	21,415,425
WT_NHS_HSF1_CS_BR2	11,366,384	2,911,272	1,781,586	6,673,526
WT-NHS_HSF1_MM_BR1	16,692,201	5,319,869	577,680	10,794,652
WT_NHS_HSF1_MM_BR2	11,242,250	3,977,681	699,178	6,565,391
***Hsf1*$^{-/-}$_NHS_HSF1_CS**	20,796,807	7,107,929	2,630,784	11,058,094
***Hsf1*$^{-/-}$_NHS_HSF1_MM**	29,863,811	14,022,020	2,229,073	13,612,718
WT_12'HS_Input	6,300,333	1,255,474	169,101	4,875,758
WT_12'HS_IgG_BR1	8,147,506	1,851,801	251,080	6,044,625
WT_12'HS_IgG_BR2	11,837,162	3,874,032	768,462	7,194,668
WT_12'HS_HSF1_CS_BR1	31,309,203	8,191,568	893,657	22,223,978
WT_12'HS_HSF1_CS_BR2	11,403,588	3,996,282	812,367	6,594,939
WT_12'HS_HSF1_MM_BR1	15,261,069	3,797,256	497,576	10,966,237
WT_12'HS_HSF1_MM_BR2	12,920,837	3,598,854	906,175	8,415,808
***Hsf1*$^{-/-}$_12'HS_HSF1_CS**	19,473,435	12,442,371	1,366,285	5,664,779
***Hsf1*$^{-/-}$_12'HS_HSF1_MM**	27,597,714	8,353,866	2,014,259	17,229,589
WT_60'HS_Input	7,767,992	1,139,060	223,600	6,405,332
WT_60'HS_IgG_BR1	14,785,033	10,619,596	224,501	3,940,936
WT_60'HS_IgG_BR2	9,523,443	2,934,886	727,287	5,861,270
WT_60'HS_HSF1_CS_BR1	13,472,052	2,369,949	427,637	10,674,466
WT_60'HS_HSF1_CS_BR2	14,863,043	3,094,488	1,093,241	10,675,314
WT_60'HS_HSF1_MM_BR1	31,268,936	23,110,624	802,871	7,355,441
WT_60'HS_HSF1_MM_BR2	12,236,596	4,055,221	636,281	7,545,094
***Hsf1*$^{-/-}$_60'HS_HSF1_CS**	23,749,525	11,390,017	2,405,648	9,953,860
***Hsf1*$^{-/-}$_60'HS_HSF1_MM**	19,278,068	9,530,952	1,604,941	8,142,175

of the HSE underneath the peaks (Figure 2.6D). While HSF1 occupies some sites prior to HS, most sites are detectably bound only after HS (Figure 2.7A) and many of these binding events appear only at 60'HS (Figure 2.7B). In terms of the genome-wide binding landscape, majority of the HSF1 peaks are located far from the nearest TSSs (Figure 2.7C). However, all the classical inducible *Hsps* have HSF1 binding within 1 kb upstream of their TSS (Figure 2.7D). Therefore, we defined a region 1kb upstream and 500 bp downstream of TSS as the promoter and examined the distribution of HSF1 binding on promoter, intragenic, and intergenic regions. The density of HSF1 peaks is highest in promoters; however, higher incidences of absolute binding events occur in intragenic and intergenic regions (Figure 2.7E).

To understand the significance of HSF1 binding in the promoter, we examined the changes in PRO-seq density in HSF1-promoter-bound genes (Figure 2.8A). 22% of HSF1-promoter-bound genes show HSF1-dependent transcription induction upon HS (magenta bars). However, 53% of HSF1-promoter-bound genes are repressed upon HS (blue bars), demonstrating that promoter bound HSF1 does not always induce transcription. Moreover, nearly all of these repressed genes are also repressed in the *Hsf1*$^{-/-}$ MEFs, indicating that promoter-bound HSF1 is not responsible for repression. Intriguingly, 13% of the HSF1-promoter-bound genes induced upon HS are also induced in *Hsf1*$^{-/-}$ MEFs (purple bar) indicating that some HSF1-bound genes do not require HSF1 for their induction.

Figure 2.7. HSF1 binding progressively increases with HS duration
(A) Venn diagram of the number of HSF1 peaks identified by MACS at different time points.
(B) Screenshot of *Des* gene, which is bound by HSF1 only after 12 min of HS, shows ChIP-seq read density in WT and *Hsf1*$^{-/-}$ MEFs. The scale on the y-axis represents the number of ChIP-seq tags under the peaks.
(C) Cumulative fraction of HSF1 peaks from the nearest annotated TSS. Colors of the cumulative distribution functions correspond to the different classes represented in the venn diagram in panel A. The dotted line represents 1 kb and the x-axis is in $\log_{10}$ scale.
(D) Cumulative fraction of the *Hsps* bound by HSF1 in the promoter and the distance of HSF1 peaks from their TSS.
(E) Density of HSF1 peaks in discrete genomic regions. The absolute numbers of HSF1 peaks are shown over the bars. Colors below the sets of three genomic regions correspond to the different classes represented in the venn diagram in panel A.

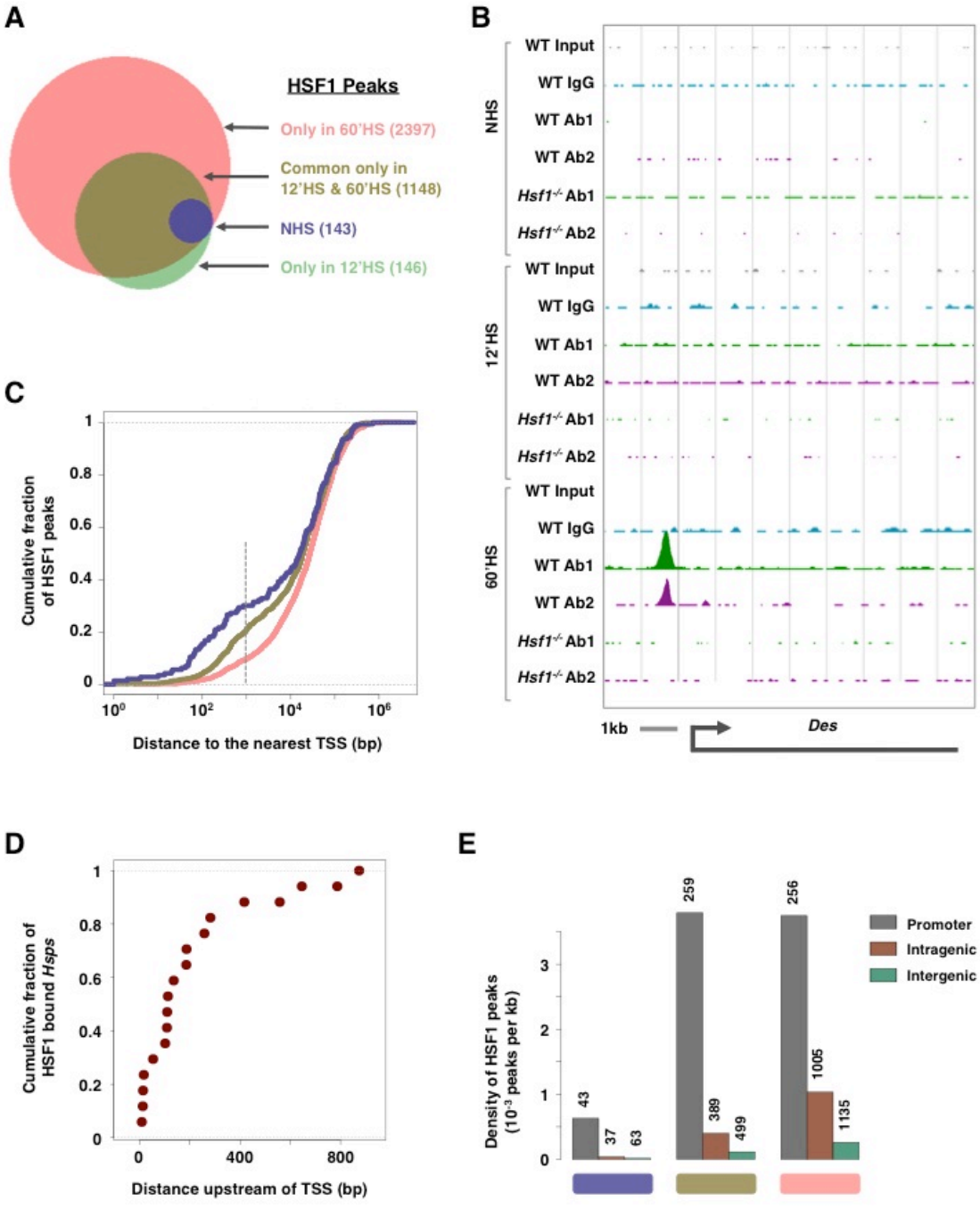

A
HSF1 Peaks
Only in 60'HS (2397)
Common only in 12'HS & 60'HS (1148)
NHS (143)
Only in 12'HS (146)
B
NHS
12'HS
60'HS
WT Input
WT IgG
WT Ab1
WT Ab2
Hsf1-/- Ab1
Hsf1-/- Ab2
1kb
Des
C
Cumulative fraction of HSF1 peaks
Distance to the nearest TSS (bp)
D
Cumulative fraction of HSF1 bound Hsps
Distance upstream of TSS (bp)
E
Density of HSF1 peaks (10-3 peaks per kb)
Promoter
Intragenic
Intergenic
43
37
63
259
389
499
256
1005
1135

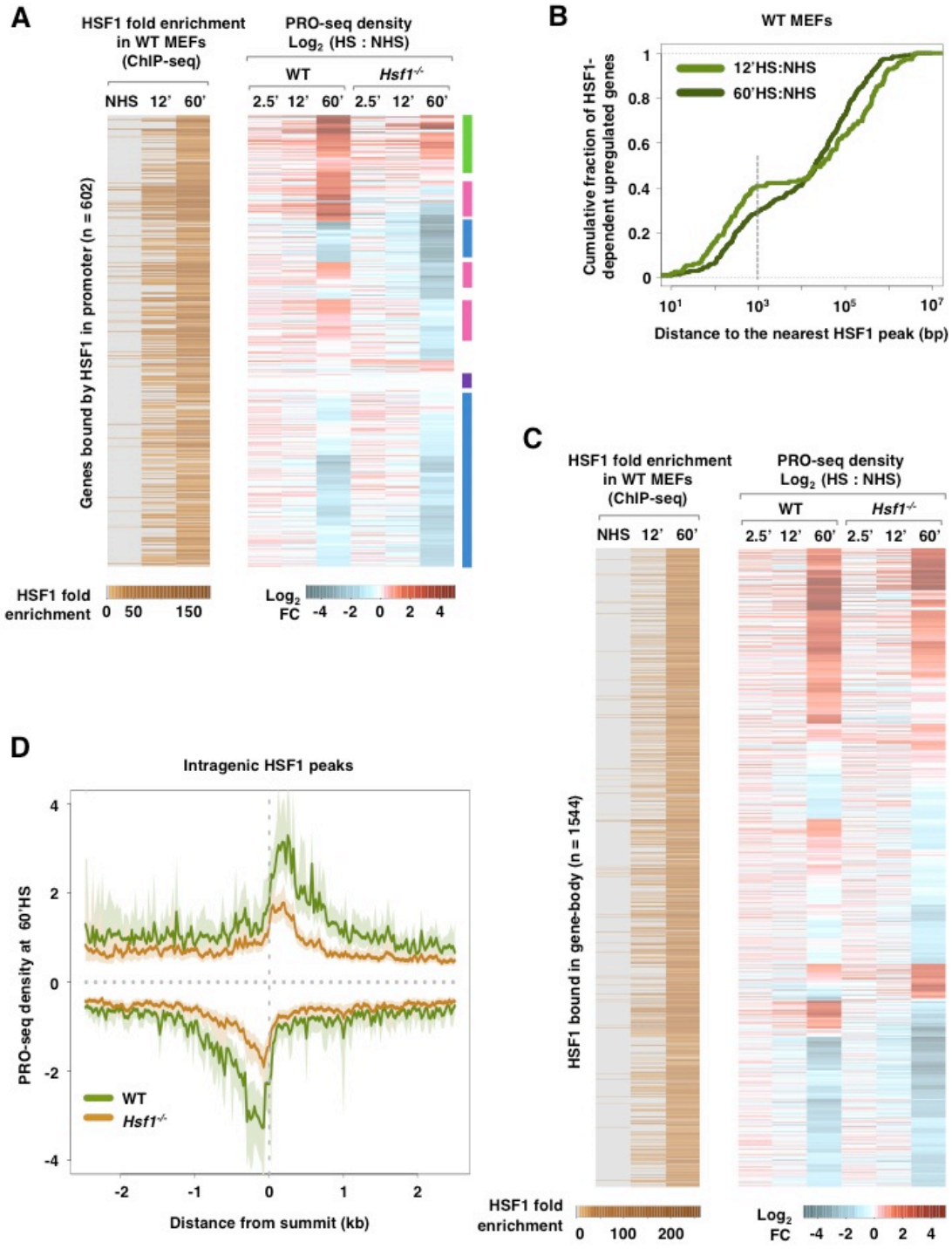
Figure 2.8 HSF1 binds and regulates a small fraction of HS induced genes

(A) Heatmap of HSF1 binding by ChIP-seq (left, grey to brown) and change in PRO-seq density (right, blue to red) before and after HS in all HSF1-promoter-bound genes. Green bar represents genes induced in both cell types, magenta bars represent HSF1-dependent induced genes, blue bars represent HSF1-independent repressed genes, and purple bar represents unchanged genes in both cell types.

(B) Cumulative fraction of HSF1-dependent upregulated genes in WT MEFs from the nearest HSF1 peaks. The dotted line represents 1 kb and x-axis is in $\log_{10}$ scale.

(C) Heatmap of HSF1 binding by ChIP-seq (left, grey to brown) and change in PRO-seq density (right, blue to red) before and after HS in all genes bound by HSF1 in the body of genes.

(D) PRO-seq density (with 95% confidence interval in light shades) at 60'HS around the center of intragenic-bound HSF1 sites. Upstream region of the intragenic-bound HSF1 sites lack HSF1-dependent accumulation of PRO-seq density.


A
HSF1 fold enrichment in WT MEFs (ChIP-seq)
PRO-seq density Log2 (HS : NHS)
WT
Hsf1-/-
NHS 12' 60'
2.5' 12' 60' 2.5' 12' 60'
Genes bound by HSF1 in promoter (n = 602)
HSF1 fold enrichment 0 50 150
Log2 FC -4 -2 0 2 4
B
WT MEFs
12'HS:NHS
60'HS:NHS
Cumulative fraction of HSF1-dependent upregulated genes
0 0.2 0.4 0.6 0.8 1
10^1 10^3 10^5 10^7
Distance to the nearest HSF1 peak (bp)
C
HSF1 fold enrichment in WT MEFs (ChIP-seq)
PRO-seq density Log2 (HS : NHS)
WT
Hsf1-/-
NHS 12' 60'
2.5' 12' 60' 2.5' 12' 60'
HSF1 bound in gene-body (n = 1544)
HSF1 fold enrichment 0 100 200
Log2 FC -4 -2 0 2 4
D
Intragenic HSF1 peaks
PRO-seq density at 60'HS
4 2 0 -2 -4
WT
Hsf1-/-
-2 -1 0 1 2
Distance from summit (kb)

We further tested whether promoter binding of HSF1 is a necessary condition for transcription induction. We probed the distance of the nearest HSF1 peak from genes that show HSF1-dependent transcription induction upon HS and found that only ~35% of these genes have HSF1 bound in their promoters (Figure 2.8B). This indicates that promoter binding of HSF1 is not entirely necessary for transcription induction during HS, and suggests that HSF1 can exert its influence from a distance, presumably from an enhancer. Together, these observations show that the promoter-bound HSF1 is not responsible for induction or repression of a majority of the HS-regulated genes.

HSF1 binding to gene-body is not the mechanism of transcription repression

Past studies have suggested that HSF1 bound in the gene-body creates an obstacle to transcribing Pol II that leads to repression of transcription (Guertin and Lis, 2010; Westwood et al., 1991). Here, we find that transcription of genes that have HSF1- bound in the transcription unit are mostly regulated in an HSF1-independent manner (Figure 2.8C). Approximately 50% of these genes are repressed upon HS; however, this repression occurs in *Hsf1*$^{-/-}$ MEFs as well. To examine the gene-body bound HSF1 mediated steric hindrance to transcribing Pol II, we examined PRO-seq density around HSF1-intragenic-bound sites at 60'HS. We detect divergent transcription at these sites, a signature of enhancers (Core et al., 2014), which

is reduced in *Hsf1*$^{-/-}$ (Figure 2.8D). PRO-seq levels upstream of the HSF1 sites reveal no significant Pol II accumulation that would be expected by the steric hindrance of bound HSF1. Thus, the HSF1 bound in the body of a gene is not a strong obstacle to transcription and does not contribute appreciably to repression during HS.

Cytoskeleton genes are induced extremely early in HSF1-independent manner

Some classical *Hsps* are known for rapid induction of transcription upon HS, for example, Drosophila *Hsp70* is induced within 2 minutes following HS (O'Brien and Lis, 1991; Zobeck et al., 2010). However, the early kinetics of induction of genes beyond *Hsps* during HS has not been examined to date. Here, we find that many genes in MEFs are significantly induced by 2.5'HS (Figure 2.9A). Majority of these early-induced (EI) genes are HSF1-independent (Figure 2.9B). Induction of these genes continues to 12'HS, after which transcription declines to below basal levels (Figure 2.9C).

Gene ontology analysis revealed that the many of these EI genes, especially the HSF1-independent ones, encode proteins with biological function related to cytoskeletal structure and function (Figure 2.9D). Dynamic rearrangement of cytoskeleton in the cells has been previously documented during HSR (Laszlo, 1992), and cytoskeleton proteins are critical for survival during heat-stress (Baird et al., 2014). Proteomics analysis also showed

Figure 2.9. Early and transiently induced genes are primarily cytoskeleton genes

(A) Screenshot of rapidly and transiently induced *Dstn* gene showing PRO-seq density at various time points in WT and *Hsf1*$^{-/-}$ MEFs.

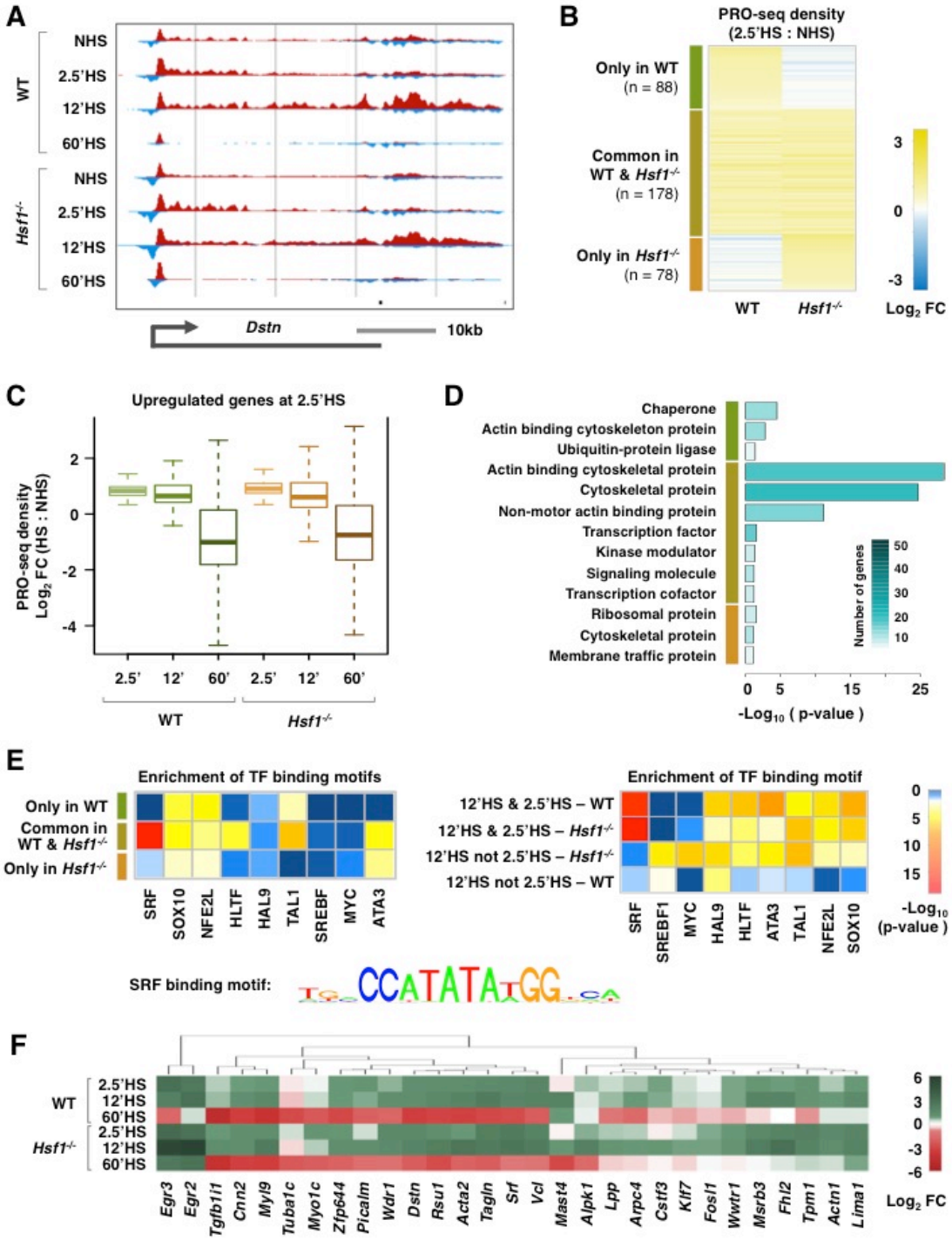
(B) Heatmap of PRO-seq density fold change in significantly upregulated genes at 2.5'HS (EI genes) in WT only (green), common in both WT & *Hsf1*$^{-/-}$ (olive), and *Hsf1*$^{-/-}$ only (orange).

(C) Change in PRO-seq density of EI genes at all time points.

(D) GO analysis of the three classes of EI genes (same color scheme as in A). The heatmap denotes the number of genes in each GO class and the length of the bar (x-axis) denotes the p-value.

(E) Significantly enriched TF motifs out of 1200 scanned in the promoter of genes in the three classes of EI genes (left panel) and the four classes of upregulated genes at 12'HS (right panel). P-value of motif enrichment over 3rd order Markov model is represented in the heatmap. Sequence logo of DNA binding motif of the most significantly enriched TF is shown.

(F) Heatmap showing PRO-seq density change during HS on a subset of immediate-early genes that were found to be bound by SRF in a previous study (Esnault et al., 2014), and contain SRF motif in promoters from our analysis.


A
WT
NHS
2.5'HS
12'HS
60'HS
Hsf1-/-
NHS
2.5'HS
12'HS
60'HS
Dstn
10kb
B
PRO-seq density
(2.5'HS : NHS)
Only in WT
(n = 88)
Common in
WT & Hsf1-/-
(n = 178)
Only in Hsf1-/-
(n = 78)
WT
Hsf1-/-
3
0
-3
Log2 FC
C
Upregulated genes at 2.5'HS
PRO-seq density
Log2 FC (HS : NHS)
2
0
-2
-4
2.5'
12'
60'
2.5'
12'
60'
WT
Hsf1-/-
D
Chaperone
Actin binding cytoskeleton protein
Ubiquitin-protein ligase
Actin binding cytoskeletal protein
Cytoskeletal protein
Non-motor actin binding protein
Transcription factor
Kinase modulator
Signaling molecule
Transcription cofactor
Ribosomal protein
Cytoskeletal protein
Membrane traffic protein
Number of genes
50
40
30
20
10
0
5
15
25
-Log10 (p-value)
E
Enrichment of TF binding motifs
Only in WT
Common in
WT & Hsf1-/-
Only in Hsf1-/-
SRF
SOX10
NFE2L
HLTF
HAL9
TAL1
SREBF
MYC
ATA3
Enrichment of TF binding motif
12'HS & 2.5'HS – WT
12'HS & 2.5'HS – Hsf1-/-
12'HS not 2.5'HS – Hsf1-/-
12'HS not 2.5'HS – WT
SRF
SREBF1
MYC
HAL9
HLTF
ATA3
TAL1
NFE2L
SOX10
0
5
10
15
-Log10
(p-value)
SRF binding motif:
F
WT
2.5'HS
12'HS
60'HS
Hsf1-/-
2.5'HS
12'HS
60'HS
Egr3
Egr2
Tgfb1i1
Cnn2
Myl9
Tuba1c
Myo1c
Zfp644
Picalm
Wdr1
Dstn
Rsu1
Acta2
Tagln
Srf
Vcl
Mast4
Alpk1
Lpp
Arpc4
Cstf3
Klf7
Fosl1
Wwtr1
Msrb3
Fhl2
Tpm1
Actn1
Lima1
6
3
0
-3
-6
Log2 FC

increase in the level of some cytoskeletal proteins in nuclear extract after two hours of HS (Raychaudhuri et al., 2014). However, the extremely rapid and transient induction of selective cytoskeleton genes during HS has not been detected before.

Because these genes are predominantly HSF1-independent, we searched for enriched TF-binding motifs in their promoters, and found serum response factor (SRF) to be the most highly enriched candidate (Figure 2.9E). SRF is known to induce a class of genes known as immediate-early genes (Schratt et al., 2001). These genes are rapidly and transiently induced in response to various extracellular stimuli (Sheng and Greenberg, 1990). SRF is also a known effector of MAP kinase pathway, which is often implicated in stress. Therefore, we examined the kinetics of induction of a subset of immediate-early genes that were found to be bound by SRF in a previous study (Esnault et al., 2014), and contain SRF motif in promoters from our analysis. Most of these genes are rapidly and transiently induced (Figure 2.9F) indicating that immediate-early genes, many of which belong to cytoskeleton family, are induced upon HS and are likely regulated by SRF.

SRF is transiently activated by HS and binds and induces cytoskeletal genes

Enrichment of DNA element for TF binding implies a possibility for regulation, however, direct evidence of binding is required to conclusively

establish the connection. To verify SRF as a novel regulator of cytoskeletal genes during HS, we examined SRF binding in WT MEFs by ChIP-seq (Figure 2.10A and Table 2.5). We found that the transiently induced genes upon HS that contain an SRF binding motif in their promoters are bound by SRF in a transient manner (Figure 2.10B). The composite profile of SRF occupancy around the TSS of all transiently induced genes also show a transient enrichment of SRF, mirroring the kinetics of transcription induction upon HS (Figure 2.10C). This finding strongly implicates SRF as a novel regulator of cytoskeletal genes during HSR.

To verify the dependence of cytoskeletal genes on SRF for transient induction during HS, we pharmacologically inhibited SRF using CCG-203971, a potent inhibitor of SRF (Bell et al., 2013; Haak et al., 2014), and made PRO-seq libraries from SRF-inhibited WT MEFs during HS (Figure 2.10D and Table 2.6). Inhibition of SRF attenuated the induction of cytoskeletal genes, as shown by the individual transcription profile of a gene (Figure 2.10E) and the metagene profile (Figure 2.10F). Together, for the first time, these finding establish SRF as a major regulator of a specific class of genes during HS.

Measurements of Pol II elongation rates show similar kinetics of induction in cytoskeleton genes and Hsps

One striking feature of PRO-seq data in induced genes is a distinct wave of elongating Pol II that are in the midst of transcription (Figure 2.11A).

Figure 2.10 SRF transiently binds and regulates transiently induced genes during HS

(A) Experimental set-up, ChIP-seq libraries in WT MEFs were made with chromatin immunoprecipitated with either non-specific IgG or SRF antibody at NHS, 12'HS, and 60'HS.

(B) Screenshot of *Vcl* gene shows the SRF ChIP-seq density (top panel) and the PRO-seq density (bottom panel) during HS. SRF binding is significantly enriched on the promoter of the gene at 12'HS and returns to the basal level by 60'HS.

(C) Composite profile of SRF ChIP-seq density around the TSS of transiently induced genes upon HS that contain SRF binding element on their promoters.

(D) Experimental set-up, PRO-seq libraries were made in duplicates at NHS and 12'HS using WT MEFs treated with either DMSO (control) or SRF inhibitor (CCG-203971).

(E) Screenshot of PRO-seq density in *Myl9* gene shows attenuation of transcription induction at 12'HS by SRF inhibitor in WT MEFs.

(F) Metagene profile of PRO-seq density at NHS and 12'HS in the scaled gene-body and a kb upstream and downstream of TSS and poly(A) site (PAS) respectively in WT MEFs treated with DMSO (top panel) and SRF inhibitor (bottom panel).

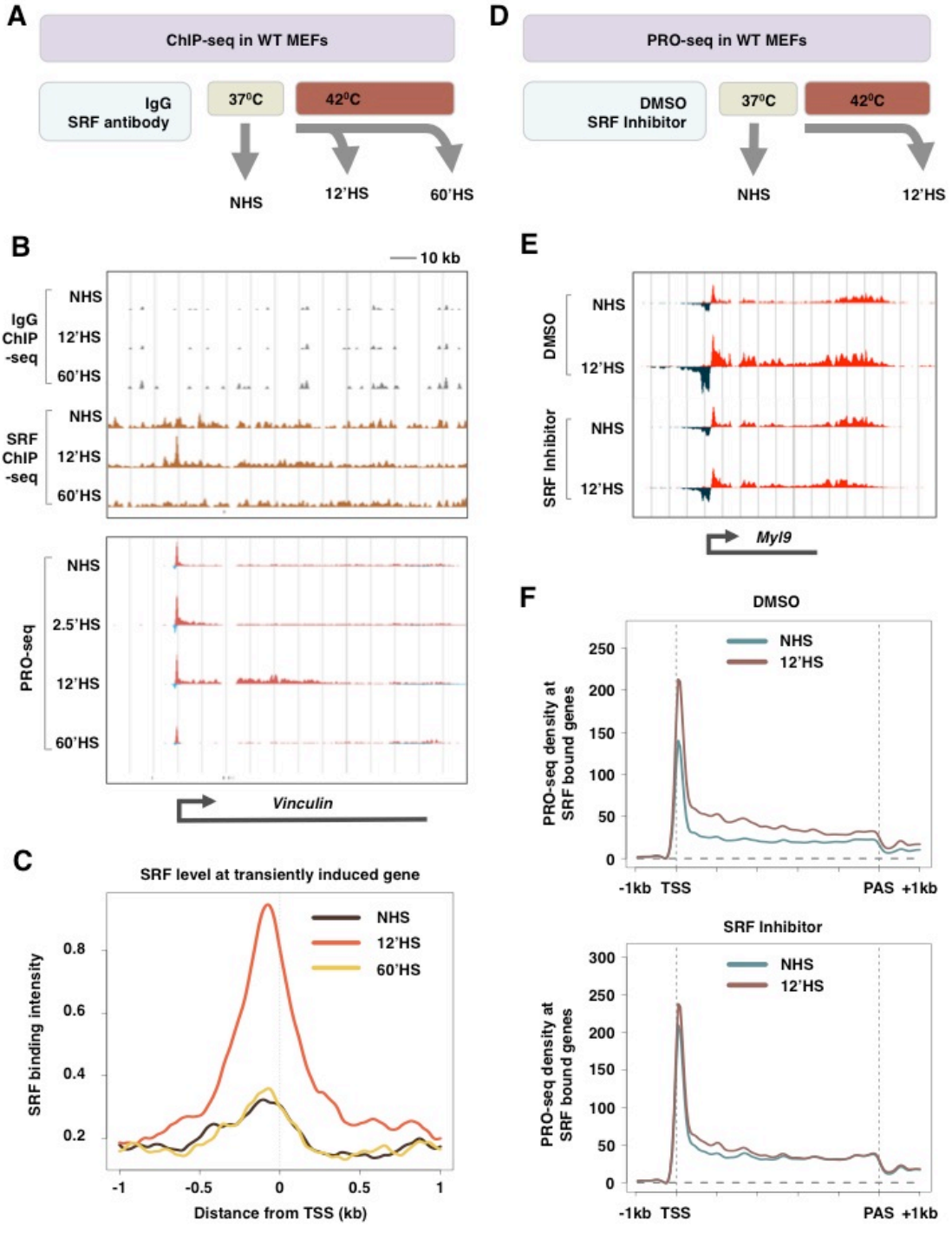

A
ChIP-seq in WT MEFs
IgG
SRF antibody
37°C
42°C
NHS
12'HS
60'HS
D
PRO-seq in WT MEFs
DMSO
SRF Inhibitor
37°C
42°C
NHS
12'HS
B
10 kb
IgG ChIP -seq
NHS
12'HS
60'HS
SRF ChIP -seq
NHS
12'HS
60'HS
PRO-seq
NHS
2.5'HS
12'HS
60'HS
Vinculin
E
DMSO
NHS
12'HS
SRF Inhibitor
NHS
12'HS
Myl9
F
DMSO
NHS
12'HS
PRO-seq density at SRF bound genes
250
200
150
100
50
0
-1kb TSS
PAS +1kb
SRF Inhibitor
NHS
12'HS
PRO-seq density at SRF bound genes
300
250
200
150
100
50
0
-1kb TSS
PAS +1kb
C
SRF level at transiently induced gene
NHS
12'HS
60'HS
SRF binding intensity
0.8
0.6
0.4
0.2
-1
-0.5
0
0.5
1
Distance from TSS (kb)

Table 2.5. Sequencing depth and alignment statistics of SRF ChIP-seq libraries. ChIP-seq reads that pass the Illumina quality filter were retained, adapters were clipped and the reads longer than 15 bp were mapped to the mouse genome.

Library	Sequenced	Adaptor dimers	Unaligned	Uniquely aligned
WT_NHS_IgG	31,175,847	3,395	611,608	30,560,844
WT_12'HS_IgG	20,594,110	4,019	445,030	20,145,061
WT_60'HS_IgG	42,710,902	2,851	1,164,766	41,543,285
WT_NHS_SRF	30,745,291	3,479	5,337,140	25,404,672
WT_12'HS_SRF	48,588,256	17,937	12,867,045	35,703,274
WT_60'HS_SRF	24,566,920	3,312	7,766,984	16,796,624

Table 2.6. Sequencing depth and alignment statistics of PRO-seq libraries in DMSO or SRF-inhibitor treated WT MEFs. Multiple aligned reads are not shown. Libraries were sequenced in Illumina HiSeq-2000.

Library	Sequenced	Adaptor dimers	Unaligned	Uniquely aligned
WT_NHS_DMSO_BR1	39,729,711	6,119,806	5,444,773	**15,181,544**
WT_NHS_DMSO_BR2	42,029,198	6,900,242	5,475,811	**16,074,055**
WT_12'HS_DMSO_BR1	2,252,435	283,488	249,187	**968,610**
WT_12'HS_DMSO_BR2	46,179,281	6,616,866	6,181,491	**18,557,328**
WT_NHS_SRFin_BR1	40,662,367	6,558,681	4,734,773	**17,152,039**
WT_NHS_SRFin_BR2	43,718,316	8,110,595	5,600,733	**16,660,386**
WT_12'HS_SRFin_BR1	43,864,184	6,146,750	5,930,696	**18,637,707**
WT_12'HS_SRFin_BR2	43,207,291	7,129,489	5,842,233	**17,229,355**

Figure 2.11. Cytoskeletal genes are induced as early as *Hsps* in HSF1-independent manner

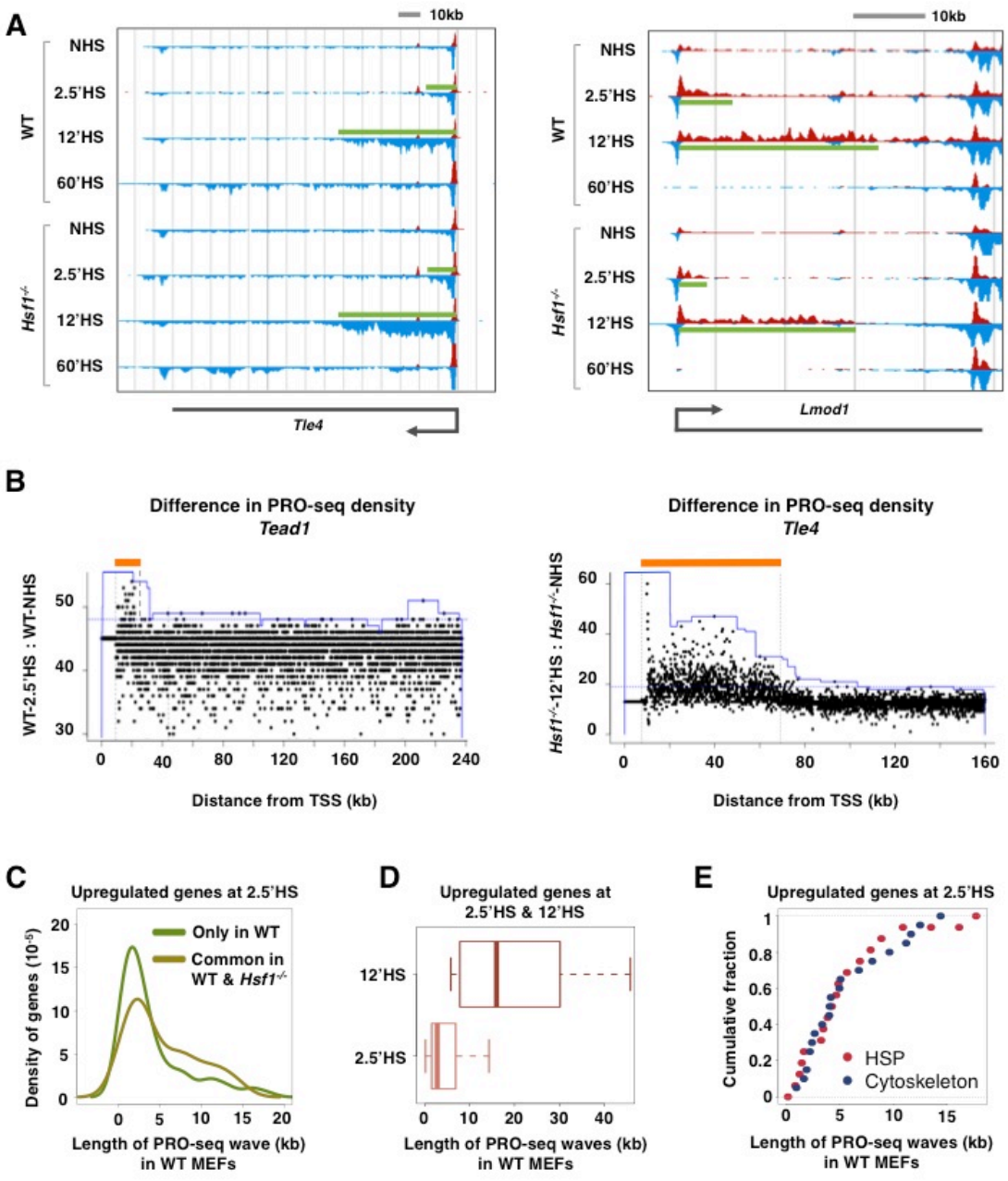
(A) Screenshots of *Tle4* and *Lmod1* genes with distinctive PRO-seq waves. Green bars represent the distance traversed by new waves of Pol II calculated by three-state HMM.

(B) Difference in PRO-seq density calculated by the three state HMM in *Tead1* gene in WT MEFs at 2.5'HS (left panel) and *Tle4* gene in *Hsf1*$^{-/-}$ MEFs at 12'HS (right panel). X-axis shows the distance from TSS, y-axis shows the difference in PRO-seq density in 50bp bins, and the orange bar represents estimated Pol II wave called by HMM.

(C) Density plot of distance traversed by PRO-seq waves in HSF1-dependent (green) and HSF1-independent (olive) genes induced at 2.5'HS.

(D) Distribution of length of PRO-seq waves at 2.5'HS and 12'HS in EI genes that are significantly upregulated at 12'HS as well.

(E) Cumulative fraction of cytoskeleton genes and *Hsps* and their length of PRO-seq waves at 2.5'HS.


A
10kb
NHS
2.5'HS
12'HS
60'HS
WT
Hsf1-/-
Tle4
Lmod1
B
Difference in PRO-seq density
Tead1
WT-2.5'HS : WT-NHS
Distance from TSS (kb)
Difference in PRO-seq density
Tle4
Hsf1-/--12'HS : Hsf1-/--NHS
Distance from TSS (kb)
C
Upregulated genes at 2.5'HS
Only in WT
Common in WT & Hsf1-/-
Density of genes (10-5)
Length of PRO-seq wave (kb) in WT MEFs
D
Upregulated genes at 2.5'HS & 12'HS
12'HS
2.5'HS
Length of PRO-seq waves (kb) in WT MEFs
E
Upregulated genes at 2.5'HS
Cumulative fraction
HSP
Cytoskeleton
Length of PRO-seq waves (kb) in WT MEFs

Here we used a three-state Hidden Markov Model (HMM) (Danko et al., 2013) to calculate the distance traveled by waves of newly released Pol II in EI genes. Our approach calculates the difference in PRO-seq density between time points in 50-bp windows throughout the gene and identifies the region (wave) with a difference in transcription (Figure 2.11B). PRO-seq wave measurement using this approach indicates that Pol II release from the pause region of EI genes occurs on average within the first minute and a half! - the median distance travelled by PRO-seq wave at 2.5'HS is 2.7 kb (Figure 2.11C), and the average elongation rate of Pol II in EI genes is 2.6 kb/min, calculated using the length of PRO-seq wave at 2.5'HS and 12'HS in genes induced at both time points (Figure 2.11D). Moreover, the length of PRO-seq waves at 2.5'HS in cytoskeleton genes and classical *Hsps* are very similar (Figure 2.11E), and the induction kinetics of HSF1-dependent and HSF1-independent genes is also highly analogous (Figure 2.11C). Together, these findings demonstrate that transcription is induced very early during HSR and the kinetics of induction is similar between HSF1-dependent (*Hsps*) and HSF1-independent (cytoskeleton) genes.

Inhibition of pause release causes massive downregulation of transcription

Historically, transcriptionally induced genes have been the focus of HS studies. However, we find that many genes are transcriptionally repressed

than induced upon HS and this global repression is independent of HSF1 (Figure 2.12A). Genes undergoing repression display two distinct kinetics: some genes are gradually and consistently repressed over the course of HS (*Pcdh18*) and others are repressed only after 12'HS (*Cdkal1*) (Figure 2.12B). The majority of the repressed genes fall into the late repressed class (Figure 2.12C).

It is unlikely that a repressor directly binds to promoters of such a large number of downregulated genes. Instead, inhibition of regulatory steps in transcription such as recruitment of Pol II or release from promoter-proximal pause seems more plausible mechanism for global repression. Consistent with this hypothesis, we see an increase in PRO-seq density in the first 100bp downstream of TSS (Figure 2.12D). This accumulation of Pol II in the 5' end of repressed genes upon HS in WT (Figure 2.13A) and *Hsf1*$^{-/-}$ MEFs (Figure 2.13B) demonstrates that the transcriptional repression in the majority of the downregulated genes is mediated by inhibition of paused Pol II release into productive elongation in an HSF1-independent manner.

To examine if the apparent extensive downregulation observed here is simply a result of thermal-induced increase in transcription rate of elongating Pol II that would decrease Pol II density over the gene-body, we calculated the elongation rate by comparing clearing wave of Pol II in downregulated genes at 12'HS and 60'HS. The average elongation rate of Pol II during HS is 2.1 kb/min (Figure 2.13C), which is highly comparable to the elongation rate of Pol

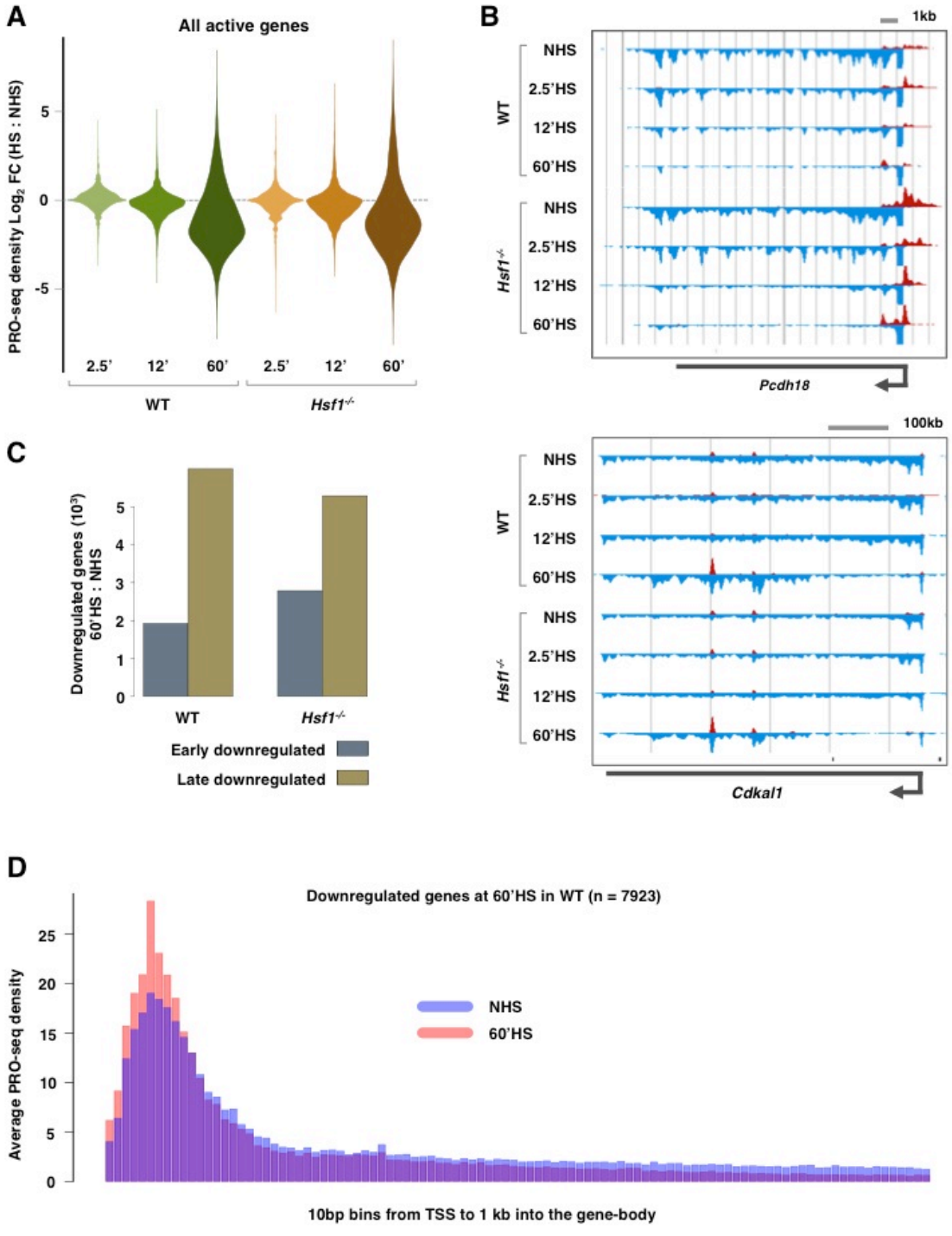
Figure 2.12. Transcription is repressed upon HS in majority of the active genes

(A) Center-of-mass plot shows the status of PRO-seq density change in all active genes (n = 15893).

(B) Screenshots of two downregulated genes with different kinetics of transcription repression upon HS: early- downregulated (top panel) late-downregulated (bottom panel).

(C) Breakdown of significantly downregulated genes at 60'HS into two classes - early and late.

(E) Histogram of average PRO-seq density in 10 bp bins from TSS to 1 kb into the gene-body for significantly downregulated genes at 60'HS (n= 7923).


A
All active genes
PRO-seq density Log2 FC (HS : NHS)
5
0
-5
2.5'
12'
60'
2.5'
12'
60'
WT
Hsf1-/-
B
1kb
WT
NHS
2.5'HS
12'HS
60'HS
Hsf1-/-
NHS
2.5'HS
12'HS
60'HS
Pcdh18
100kb
WT
NHS
2.5'HS
12'HS
60'HS
Hsf1-/-
NHS
2.5'HS
12'HS
60'HS
Cdkal1
C
Downregulated genes (10^3)
60'HS : NHS
5
4
3
2
1
0
WT
Hsf1-/-
Early downregulated
Late downregulated
D
Downregulated genes at 60'HS in WT (n = 7923)
Average PRO-seq density
25
20
15
10
5
0
NHS
60'HS
10bp bins from TSS to 1 kb into the gene-body

II under normal temperature in mouse embryonic stem cells (1.8 – 2.4 kb/min)(Jonkers et al., 2014). Furthermore, the downregulation of transcription observed by PRO-seq is nicely reflected at the mRNA level measured by RNA-seq (Figure 2.4B, right panel). Thus, the decrease in gene-body PRO-seq reads is not the result of an increase in elongation rate, but rather results from a genuine downregulation in the frequency with which Pol II transitions from the pause to productive elongation.

We performed GO analyses to examine the enriched functional classes in downregulated genes. Genes involved in metabolism, cell cycle, and mitosis are represented in the early-repressed class, while genes in the late-repressed class are enriched for mRNA splicing, mRNA processing and nuclear transport functions (Figure 2.13D). Conceivably, cells may enter into a low metabolic profile upon HS by immediately shutting down genes involved in cell cycle and metabolism. However, mRNA processing and efficient splicing could be critical in the early phase of HSR when *Hsps* and IE genes are robustly induced. These two kinetic classes of genes that have distinct functions imply that repression of transcription is a highly-regulated process, rather than a previously suggested non-discriminatory HS-induced global repression of transcription (Teves and Henikoff, 2011).

HS-regulated Hsps are dependent on HSF1

Expression of *Hsps* has served as an indicator of HSR. There are ~75

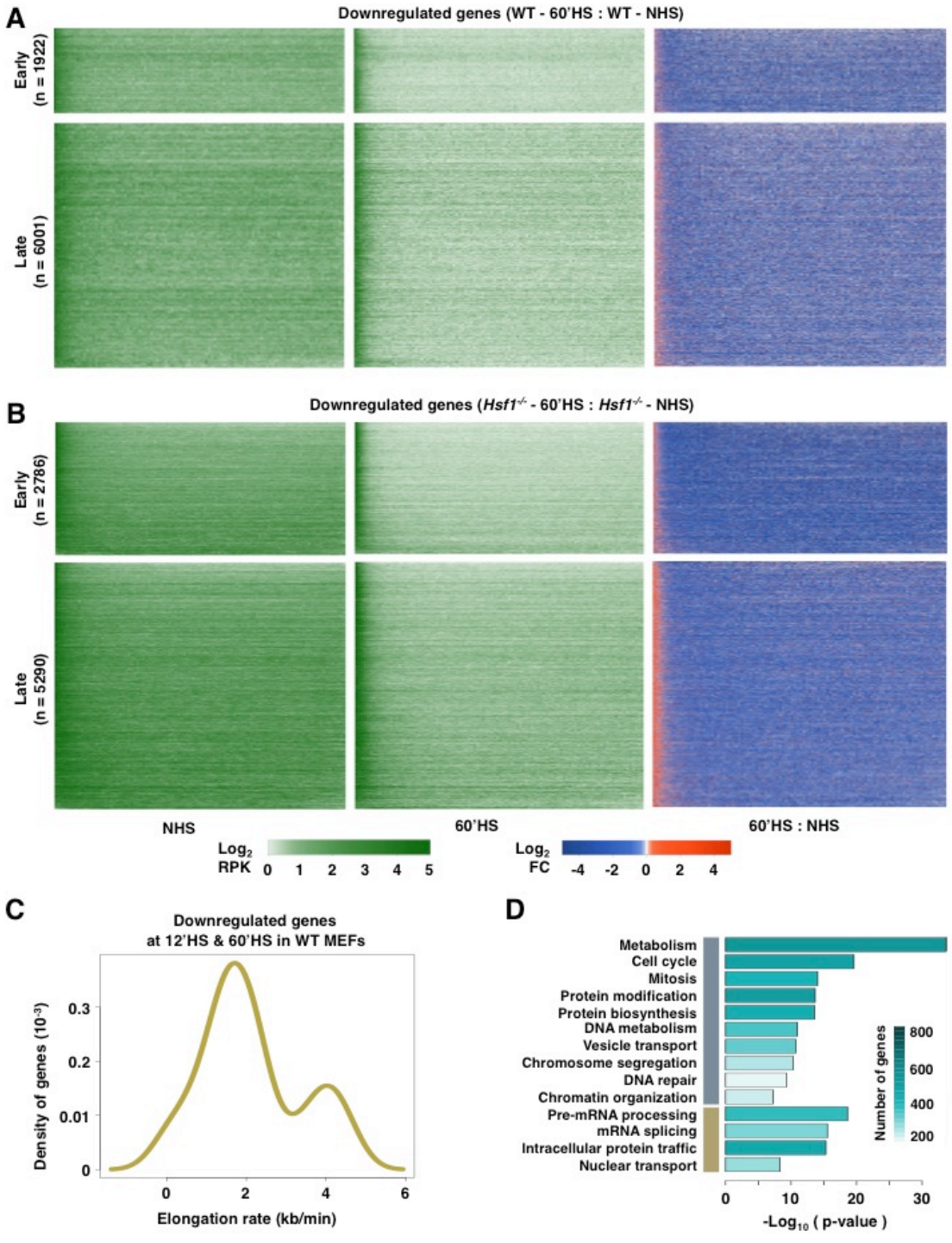
Figure 2.13. Global downregulation during HS is mediated by lack of paused Pol II release

(A) Heatmap of PRO-seq density before HS (left), after 60'HS (mid), and fold change from 60'HS to NHS (right) for significantly downregulated genes (early – top panels and late – bottom panels) at 60'HS in WT MEFs. Each row represents a gene, scaled to same length and divided into 100 bins, from TSS up to polyA site for genes shorter than 24 kb and up to 24 kb for genes longer than 24kb.

(B) Heatmap of PRO-seq density before HS (left), after 60'HS (mid), and fold change from 60'HS to NHS (right) for significantly downregulated genes (early – top panels and late – bottom panels) at 60'HS in *Hsf1*$^{-/-}$ MEFs. Each row represents a gene, scaled to same length and divided into 100 bins, from TSS up to polyA site for genes shorter than 24 kb and up to 24 kb for genes longer than 24kb.

(C) Density of elongation rates of Pol II during HS in downregulated genes. Elongation rates were calculated by first measuring the distance travelled by Pol II between 12'HS and 60'HS in significantly downregulated genes at both 12'HS and 60'HS and then dividing the distance by the time (60 - 12 = 48 minutes).

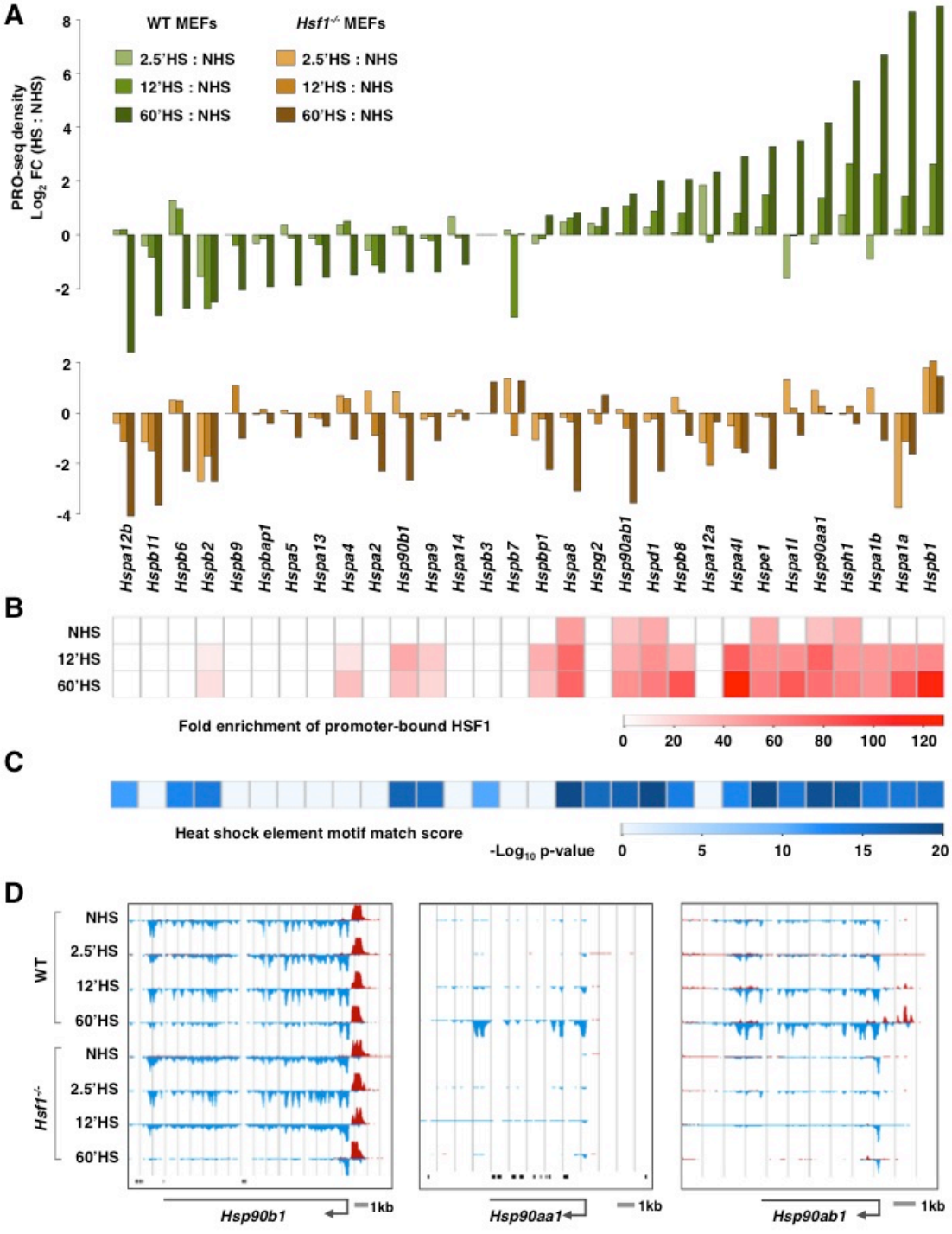
(D) GO analysis of the significantly downregulated genes using DAVID. GO terms enriched in early and late downregulated classes are represented by color scheme in Figure 2.12C. The heatmap denotes the number of genes in each GO class.


A
Downregulated genes (WT - 60'HS : WT - NHS)
Early (n = 1922)
Late (n = 6001)
B
Downregulated genes (Hsf1-/- - 60'HS : Hsf1-/- - NHS)
Early (n = 2786)
Late (n = 5290)
NHS
60'HS
60'HS : NHS
Log2 RPK 0 1 2 3 4 5
Log2 FC -4 -2 0 2 4
C
Downregulated genes at 12'HS & 60'HS in WT MEFs
Density of genes (10-3)
0.3
0.2
0.01
0
0 2 4 6
Elongation rate (kb/min)
D
Metabolism
Cell cycle
Mitosis
Protein modification
Protein biosynthesis
DNA metabolism
Vesicle transport
Chromosome segregation
DNA repair
Chromatin organization
Pre-mRNA processing
mRNA splicing
Intracellular protein traffic
Nuclear transport
Number of genes
800
600
400
200
0 10 20 30
-Log10 (p-value)

genes in HSP family (Kampinga et al., 2009), some of which are highly related paralogs and likely serve in different cellular compartments, cell types, or developmental stages. It is unclear whether transcription of all *Hsps* increases upon HS. Here, we examined transcriptional change in genes in the HSP family. While classical *Hsps* show HSF1-dependent induction upon HS, more than half of *Hsps* are transcriptionally repressed (Figures 2.14A and 2.15A). For example, *Hsp90b1*, an endoplasmic reticulum associated chaperone involved in unfolded protein response, is repressed upon HS in both cell types, while other *Hsp90s* are robustly induced in an HSF1-dependent manner (Figure 2.14D). Strikingly, almost all transcriptionally induced *Hsps* have prominent HSF1 peaks in their promoters (Figures 2.14B and 2.15B), accompanied by the presence of HSEs underneath the peaks (Figures 2.14C and 2.15C). Overall, our results show that some *Hsps* are rapidly and robustly induced upon HS; however, many genes in the HSP family are not bound by HSF1 and are not transcriptionally induced in response to HS, suggesting the role of post-transcriptional (Theodorakis and Morimoto, 1987; Yost et al., 1990) and translational regulation (Lindquist, 1987; Storti et al., 1980; Zhou et al., 2015) in increase in gene expression of some *Hsps* upon HS or the likely specialization of different HSPs in different tissues or stress responses.

HSF1 induces transcription by increasing promoter-proximal pause release

Figure 2.14. HSF1 is required for induction of classical *Hsps* upon HS
(A) PRO-seq density change upon HS in 30 *Hsp* genes in WT MEFs (top panel) and *Hsf1*$^{-/-}$ MEFs (bottom panel) ordered by increasing fold change at WT 60'HS.
(B) HSF1 fold enrichment in the promoter of the corresponding 30 *Hsp* genes in WT MEFs before and after HS.
(C) Motif match score p-value of HSE in the promoter of the corresponding 30 *Hsp* genes.
(D) Screenshots of three *Hsp90* genes. *HSP90b1* is downregulated upon HS in HSF1-independent manner and *Hsp90aa1* and *Hsp90ab1* are robustly upregulated in HSF1-dependent manner.


A
WT MEFs
Hsf1-/- MEFs
2.5'HS : NHS
12'HS : NHS
60'HS : NHS
2.5'HS : NHS
12'HS : NHS
60'HS : NHS
PRO-seq density
Log2 FC (HS : NHS)
8
6
4
2
0
-2
2
0
-2
-4
Hspa12b
Hspb11
Hspb6
Hspb2
Hspb9
Hspbap1
Hspa5
Hspa13
Hspa4
Hspa2
Hsp90b1
Hspa9
Hspa14
Hspb3
Hspb7
Hspbp1
Hspa8
Hspg2
Hsp90ab1
Hspd1
Hspb8
Hspa12a
Hspa4l
Hspe1
Hspa1l
Hsp90aa1
Hsph1
Hspa1b
Hspa1a
Hspb1
B
NHS
12'HS
60'HS
Fold enrichment of promoter-bound HSF1
0
20
40
60
80
100
120
C
Heat shock element motif match score
-Log10 p-value
0
5
10
15
20
D
WT
Hsf1-/-
NHS
2.5'HS
12'HS
60'HS
NHS
2.5'HS
12'HS
60'HS
Hsp90b1
1kb
Hsp90aa1
1kb
Hsp90ab1
1kb

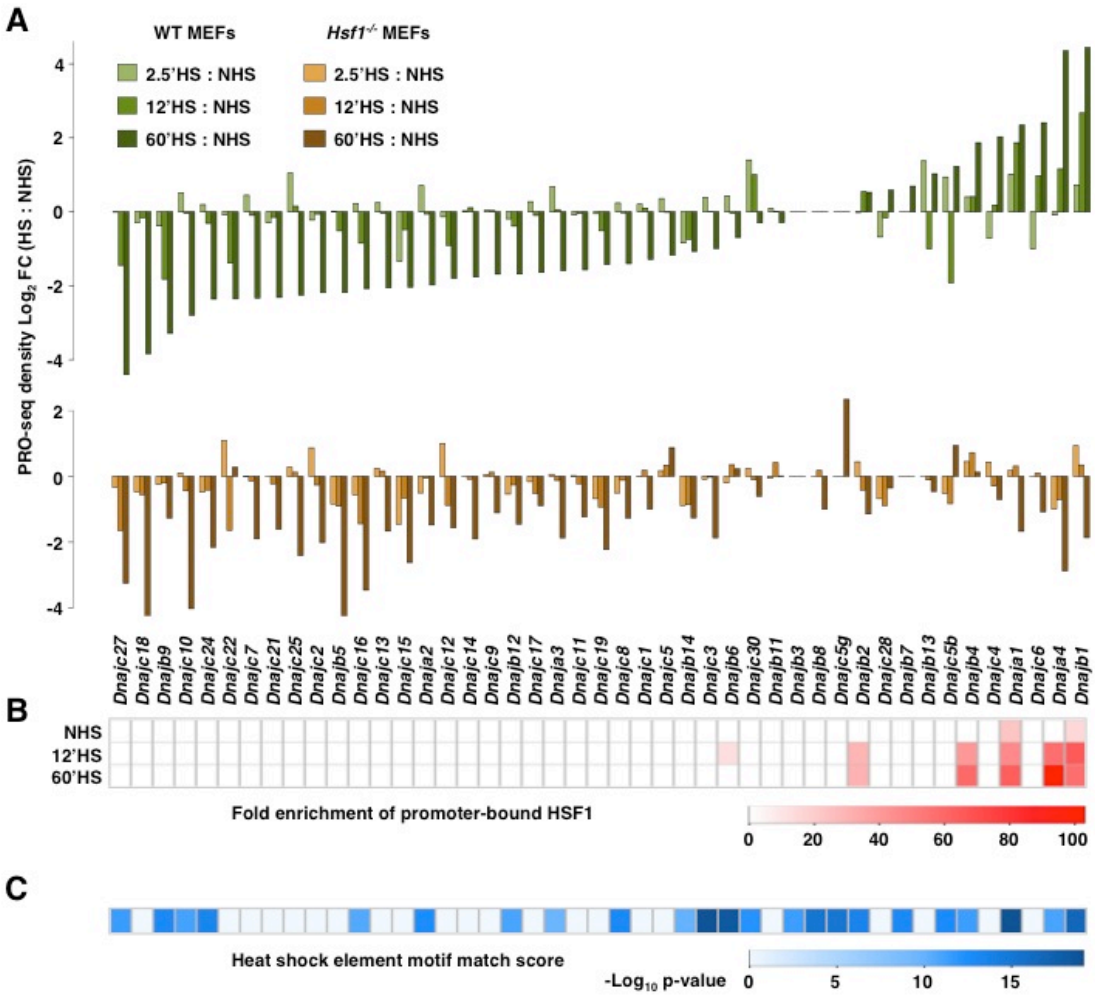


Figure 2.15. Most *Hsp40s* are not induced by HS
(A) PRO-seq density change upon HS in 45 *Hsp40* genes in WT MEFs (top panel) and *Hsf1*$^{-/-}$ MEFs (bottom panel) ordered by increasing fold change at WT 60'HS.
(B) HSF1 fold enrichment in the promoter of the corresponding 45 *Hsp40* genes in WT MEFs before and after HS.
(C) Motif match score p-value of HSE in the promoter of the corresponding 45 *Hsp40* genes.

The PRO-seq profiles across the promoter proximal and gene-body regions provide insight to Pol II's progress through the distinct steps in the transcription cycle and the changes occurring in response to HS. Moreover, contrasting these profiles in both WT and *Hsf1*$^{-/-}$ MEFs reveals the role of HSF1 in the changes observed. We first examined the genes that exhibit a significant increase in PRO-seq density in their gene-body by 60'HS in WT but not in *Hsf1*$^{-/-}$ MEFs (n=102) (Figure 2.16A, middle panel) and also have HSF1 bound in their promoter region (Figure 2.16A, left panel). A simple interpretation of these profiles is that HSF1 acts to increase the rate of Pol II release from the pause, a step that is accelerated by P-TEFb kinase, which is known to be recruited to HS loci in an HSF1-dependent manner (Lis et al., 2000). P-TEFb phosphorylates components of the paused Pol II complex and enables Pol II to embark into productive elongation (Renner et al., 2001). An alternative mechanism to explain these profiles is that HS-induces an HSF1-dependent decrease in early termination that leads to higher gene-body Pol II density. However, this anti-termination model was ruled out, at least for *Hsp70*, by our previous study showing the termination rate of promoter-proximal paused Pol II is similar in HS and NHS cells and could not account for the HS-induced increase in gene-body Pol II density (Buckley et al., 2014). A second alternative explanation for these profiles is that elongation rates are slowed during HS in an HSF1-dependent manner leading to increases in gene-body Pol II density. However, such a decrease in elongation rate is

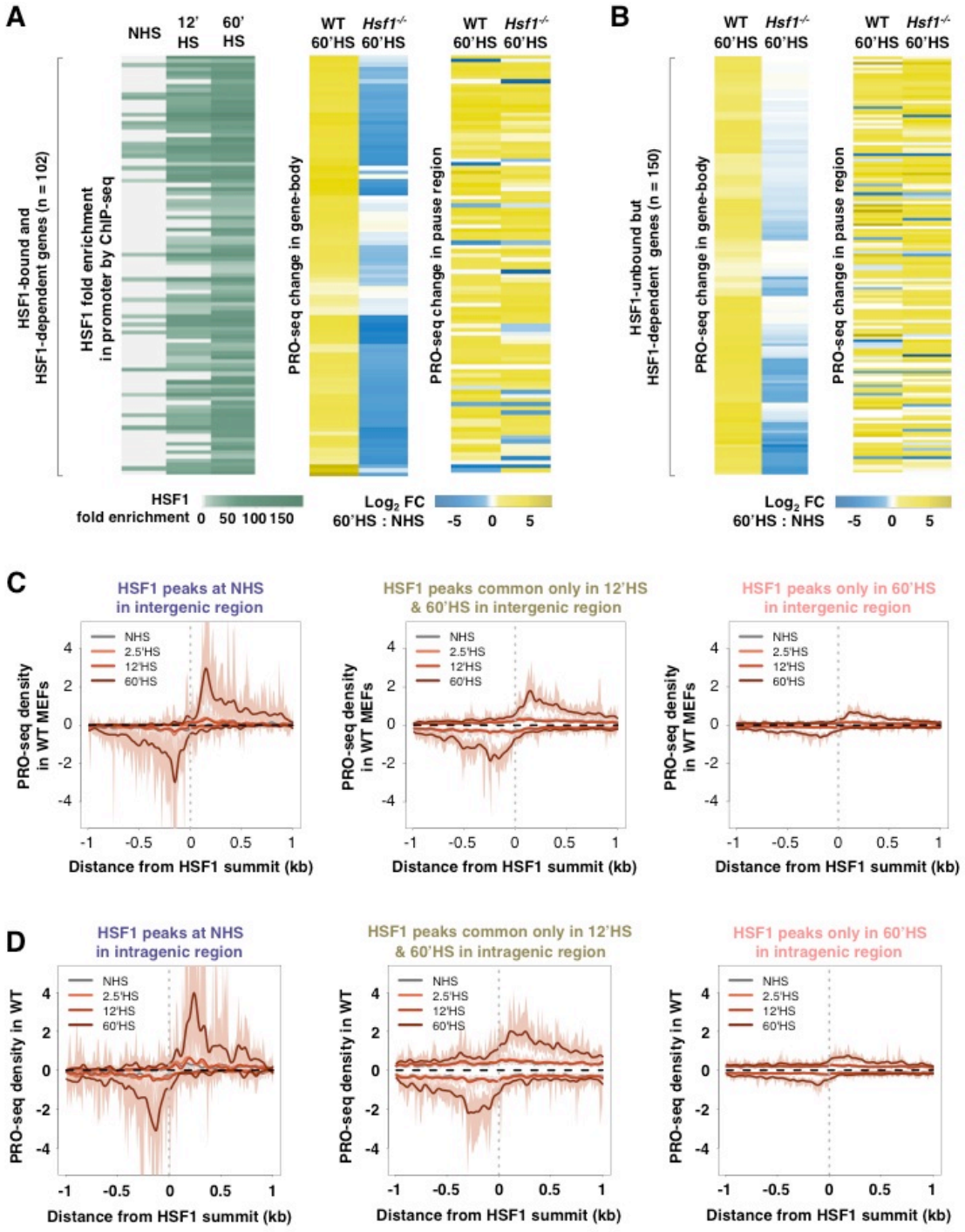
Figure 2.16. HSF1 induces transcription by increasing paused Pol II release

(A) Heatmap of HSF1 binding (left), change in PRO-seq density in gene-body (center) and pause region (right) of the 102 genes that are bound by HSF1 in the promoter and show HSF1-dependent transcription induction at 60'HS.

(B) Change in PRO-seq density in gene-body (left) and pause region (right) of the 150 genes that are not bound by HSF1 in the promoter but show HSF1-dependent transcription induction at 60'HS.

(C) PRO-seq density (with 95% confidence interval in light shades) before and at time points after HS in WT MEFs around the center of HSF1-intergenic-bound sites at NHS (left), common only in 12'HS & 60'HS (center), and only in 60'HS (right). The color of text in the titles refers to the three classes represented by the same colors in Figure 2.7A.

(D) PRO-seq density (with 95% confidence interval in light shades) before and at time points after HS in WT MEFs around the center of HSF1-intragenic-bound sites at NHS (left), common only in 12'HS & 60'HS (center), and only in 60'HS (right). The color of text in the titles refers to the three classes represented by the same colors in Figure 2.7A.


A
NHS
12' HS
60' HS
WT 60'HS
Hsf1-/- 60'HS
WT 60'HS
Hsf1-/- 60'HS
HSF1-bound and HSF1-dependent genes (n = 102)
HSF1 fold enrichment in promoter by ChIP-seq
PRO-seq change in gene-body
PRO-seq change in pause region
HSF1 fold enrichment 0 50 100 150
Log2 FC 60'HS : NHS -5 0 5
B
WT 60'HS
Hsf1-/- 60'HS
WT 60'HS
Hsf1-/- 60'HS
HSF1-unbound but HSF1-dependent genes (n = 150)
PRO-seq change in gene-body
PRO-seq change in pause region
Log2 FC 60'HS : NHS -5 0 5
C
HSF1 peaks at NHS in intergenic region
HSF1 peaks common only in 12'HS & 60'HS in intergenic region
HSF1 peaks only in 60'HS in intergenic region
NHS
2.5'HS
12'HS
60'HS
PRO-seq density in WT MEFs
Distance from HSF1 summit (kb)
D
HSF1 peaks at NHS in intragenic region
HSF1 peaks common only in 12'HS & 60'HS in intragenic region
HSF1 peaks only in 60'HS in intragenic region
NHS
2.5'HS
12'HS
60'HS
PRO-seq density in WT
Distance from HSF1 summit (kb)

neither consistent with the increases in the mRNA levels of these genes (Figures 2.4B and 2.4C), nor supported by the direct measurement of elongation rates (Figures 2.11C-E and 2.13C) (Ardehali and Lis, 2009). Thus, these 102 HSF1-promoter-bound and transcriptionally induced genes appear to be regulated by HSF1 upon HS in an HSF1-dependent manner at the step of pause Pol II release into productive elongation.

Interestingly, this class of 102 genes also shows an increase in PRO-seq density in their pause region upon HS. In contrast to the gene-body, this increase is independent of HSF1 (Figure 2.16A, right panel). We hypothesize that this increase is simply a consequence of the massive HSF1-independent downregulation of genes (Figures 2.4A & 2.4D) upon HS, which increases the cellular pool of Pol II and thereby drives Pol II loading on promoters by mass action in the presence and absence of HSF1. Furthermore, this model is consistent with the fact that the broad downregulation of thousands of genes is also HSF1 independent (Figure 2.4D). This hypothesis can be tested by quantifying chromatin-bound and chromatin-unbound Pol II during HS.

HSF1 is also capable of acting from a distal enhancer, at least when bound to sites composed of an array of HSEs (Bienz and Pelham, 1986). We identified many genes that show HSF1-dependent induction upon HS but don't have HSF1 bound in their promoters (n=150) (Figure 2.8B) suggesting that these genes could be regulated by HSF1 acting from distal enhancers. Indeed, these genes show the same changes in PRO-seq profile in both gene bodies

and pause regions as the class of 102 HSF1-promoter-bound, HSF1-dependent induced genes (Figure 2.16B). Therefore, HSF1 bound at distal enhancers likely regulates genes by increasing the release of paused Pol II, indicating that the promoter-proximal pause release is the primary step regulated in all HSF1-mediated transcription induction during HSR.

HSF1 enhancers would be expected to have the divergent transcription profile characteristic of active enhancers (Core et al., 2014; Hah et al., 2011; Kim et al., 2010). PRO-seq profiles around the intergenic and intragenic HSF1-bound regions show that the divergent transcription appears upon HS (Figures 2.16C and 2.16D), suggesting that these distal HSF1 bound regions become active enhancers during the HSR. This increase in divergent transcription around the distal HSF1-bound sites is unlikely to be an effect of HSF1 directly stimulating the recruitment of Pol II, because the initiation of divergent transcription is detectable only at 60'HS even though the sites are occupied by HSF1 prior to 60'HS (Figures 2.16C and 2.16D - left and mid panels). The delayed recruitment of Pol II is more consistent with an indirect role of HSF1 on Pol II recruitment to enhancers, perhaps a consequence of opening of promoter regions that allows the binding of other TFs and Pol II, and of the increased pool of Pol II made available during the massive downregulation of thousands of genes.

HS-regulated genes have concomitant change in P-TEFb level

Our findings implicate the step of Pol II release from the promoter-proximal pause as the major regulatory knob in transcription cycle during HS. HSF1 induces transcription by promoting the release of paused Pol II and the massive repression is caused by the lack of efficient release of paused Pol II. These observations encouraged us to probe the change in global occupancy of the pause release factor P-TEFb during HS. P-TEFb has two subunits – cyclin T and cyclin-dependent kinase 9 (Cdk9) (Zhou et al., 2012). Because the Pol II pause release is dependent on the kinase subunit (Peterlin and Price, 2006), we performed ChIP-seq before and after 60 min of HS in WT MEFs using Cdk9 antibody (Figure 2.17A and Table 2.7).

The occupancy of Cdk9 significantly increases upon HS in the representative *Hsph1* gene (Figure 2.17B) and all HS-inducible *Hsps* (Figure 2.17C), whereas the Cdk9 level in HS-uninducible *Hsps* remain unchanged (Figure 2.17D). Cdk9 occupancy also increases in all HSF1-dependent upregulated genes that include a much larger pool of genes beyond HS-inducible *Hsps* (Figure 2.17E), suggesting interplay between HSF1 and P-TEFb in transcription induction by increasing Pol II pause release. Although the pause region has more prominent increase in P-TEFb level, this increase occurs throughout the gene body indicating that P-TEFb tracks with Pol II into gene body (Jonkers and Lis, 2015; Ni et al., 2008). Similarly, the decrease in Pol II release from promoter-proximal pause of transcriptionally repressed genes during HS suggests a decrease in P-TEFb activity. Therefore, we

Figure 2.17. P-TEFb is enriched in induced and depleted in repressed genes

(A) Experimental set-up, ChIP-seq libraries in WT MEFs were made with chromatin immunoprecipitated with either non-specific IgG or Cdk9 antibody at NHS and 60'HS.

(B) Screenshot of *Hsph1* gene shows the ChIP-seq density in WT MEFs with IgG or Cdk9 antibody during HS. Cdk9 occupancy is significantly enriched on the promoter gene-body at 60'HS.

(C) Metagene profile of normalized Cdk9 density in HS-inducible *Hsps* at NHS and 60'HS in WT MEFs. The x-axis represents the scaled gene-body and a kb upstream and downstream of TSS and PAS respectively.

(D) Metagene profile of normalized Cdk9 density in HS-uninducible *Hsps* at NHS and 60'HS in WT MEFs. The x-axis represents the scaled gene-body and a kb upstream and downstream of TSS and PAS respectively.

(E) Metagene profile of normalized Cdk9 density in all HSF1 bound and HS-inducible genes at NHS and 60'HS in WT MEFs. The x-axis represents the scaled gene-body and a kb upstream and downstream of TSS and PAS respectively.

(F) Metagene profile of normalized Cdk9 density around the TSS of HS-repressed genes at NHS and 60'HS in WT MEFs.

(G) Screenshots of *Cdk9* (top panel) and *Ccnt1* (bottom panel) genes, two subunits of P-TEFb complex, showing a lack of transcriptional regulation during HS.

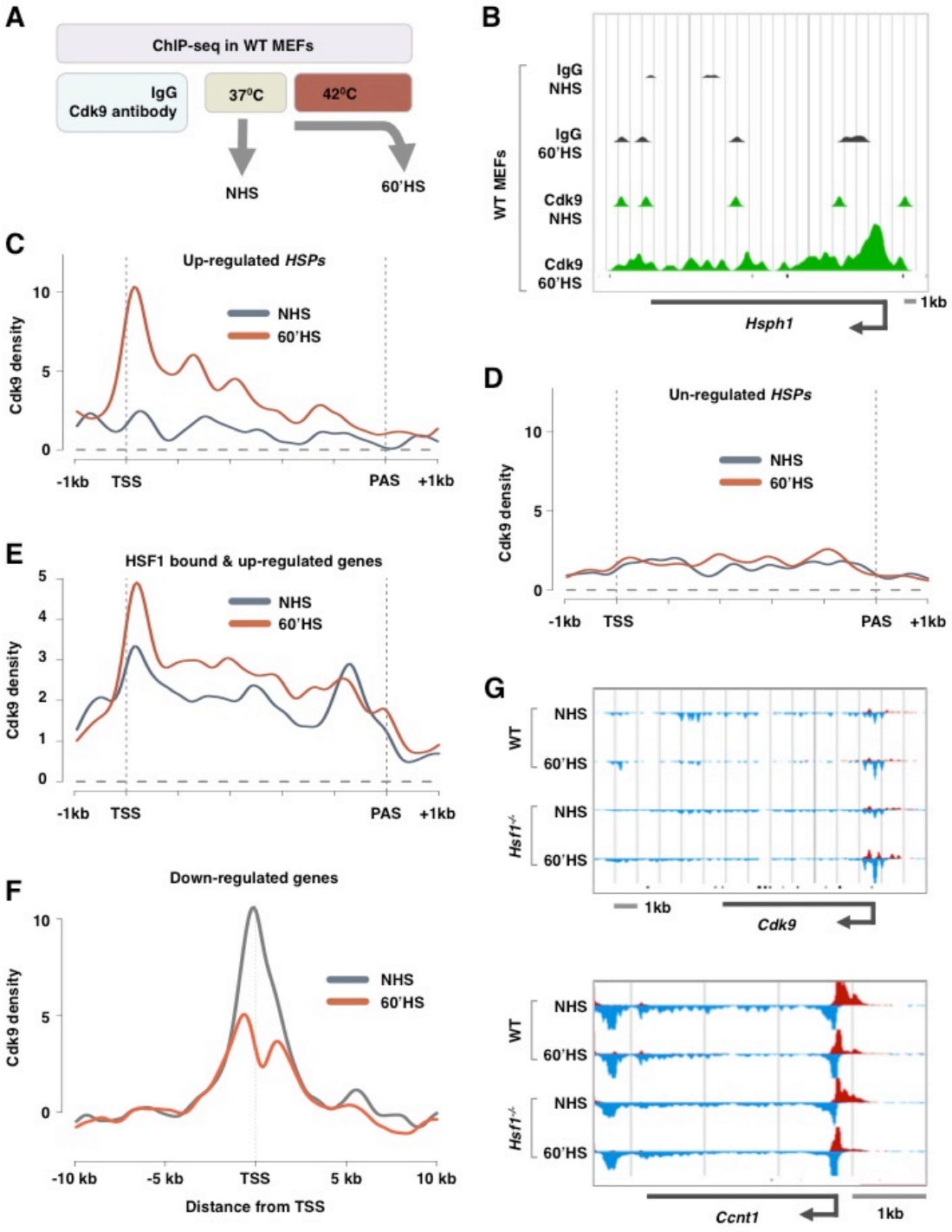

A
ChIP-seq in WT MEFs
IgG
Cdk9 antibody
37°C
42°C
NHS
60'HS
B
IgG
NHS
IgG
60'HS
Cdk9
NHS
Cdk9
60'HS
WT MEFs
1kb
Hsph1
C
Up-regulated HSPs
NHS
60'HS
Cdk9 density
10
5
0
-1kb
TSS
PAS
+1kb
D
Un-regulated HSPs
NHS
60'HS
Cdk9 density
10
5
0
-1kb
TSS
PAS
+1kb
E
HSF1 bound & up-regulated genes
NHS
60'HS
Cdk9 density
5
4
3
2
1
0
-1kb
TSS
PAS
+1kb
F
Down-regulated genes
NHS
60'HS
Cdk9 density
10
5
0
-10 kb
-5 kb
TSS
5 kb
10 kb
Distance from TSS
G
WT
NHS
60'HS
Hsf1-/-
NHS
60'HS
1kb
Cdk9
WT
NHS
60'HS
Hsf1-/-
NHS
60'HS
Ccnt1
1kb

Table 2.7. Sequencing depth and alignment statistics of Cdk9 ChIP-seq libraries. ChIP-seq reads that pass the Illumina quality filter were retained, adapters were clipped and the reads longer than 15 bp were mapped to the mouse genome.

Library	Sequenced	Adaptor dimers	Unaligned	Uniquely aligned
WT_NHS_IgG	31,175,847	3,395	611,608	30,560,844
WT_12'HS_IgG	20,594,110	4,019	445,030	20,145,061
WT_60'HS_IgG	42,710,902	2,851	1,164,766	41,543,285
WT_NHS_Cdk9	7,876,051	1,525	1,010,967	6,863,559
WT_12'HS_Cdk9	21,852,365	1,929	2,145,718	19,704,718
WT_60'HS_Cdk9	24,589,485	6,106	3,055,814	21,527,565

quantified Cdk9 level around the TSS of HS-repressed genes and as expected found a reduction in Cdk9 occupancy (Figure 2.17F).

Together, these observations suggest that the increase and decrease in transcription level in different sets of genes is tightly coordinated - an increase in free pool of nuclear P-TEFb level due to global downregulation of genes during HS paves the groundwork for HSF1-dependent increase in P-TEFb level at the HS-induced genes, likely through the known HSF1-dependent recruitment of P-TEFb kinase (Lis et al., 2000). The availability of free pool of P-TEF-b level is critical for transcription induction of survival genes because the transcription of either subunits of P-TEFb is uninduced during HS (Figure 2.17G) and the nuclear P-TEFb level decreases during HS (Raychaudhuri et al., 2014). Thus, the measurement of global Cdk9 occupancy change implicates a critical role for P-TEFb in both upregulation and downregulation of transcription during HS.

HS-regulated genes have different kinetics, dynamics, chromatin marks, and functions

Our results demonstrate an elaborate network of transcriptional regulation in response to heat stress; hundreds of genes are induced and thousands of others are repressed. This coordinated regulation exhibits temporal precision and selectivity for functional gene groups. We broadly

divided the kinetics and dynamics of transcription regulation during HS in five unique classes (Figure 2.18).

Class I represents the HS-inducible genes including *Hsps* that are induced throughout the duration of HS in this study. Upon HS, many of these genes are bound by HSF1 to the promoter. The binding motif of AP-1 is enriched within the HSF1 peaks in the promoter of these genes (Figure 2.19A) suggesting a likely co-operation between HSF1 and AP-1. We show that the HSF1 induces transcription in HSF1-dependent genes of this class by increasing the release of paused Pol II into productive elongation. In addition to HSF1, the binding motif of SIX4 is also enriched in the promoters of some genes in this class (Figure 2.19B).

Class II comprises EI genes described in Figure 3. HS rapidly and transiently induces transcription of these genes. Many genes in this class belong to cytoskeleton family and are characterized by a strong enrichment of SRF binding motif in their promoters. Metagene profile of PRO-seq density in the pause region of these genes also shows a comparable net increase of paused Pol II at 2.5'HS and 12'HS that decrease at 60'HS (Figure 2.19C). Thus, both Pol II pause release and recruitment appear rapidly and transiently regulated during HSR, and SRF and its coregulators are strong candidates for mediating this response.

Class III represents genes that are induced during the late phase of the HSR. Most of these genes are regulated independently of HSF1 and are likely

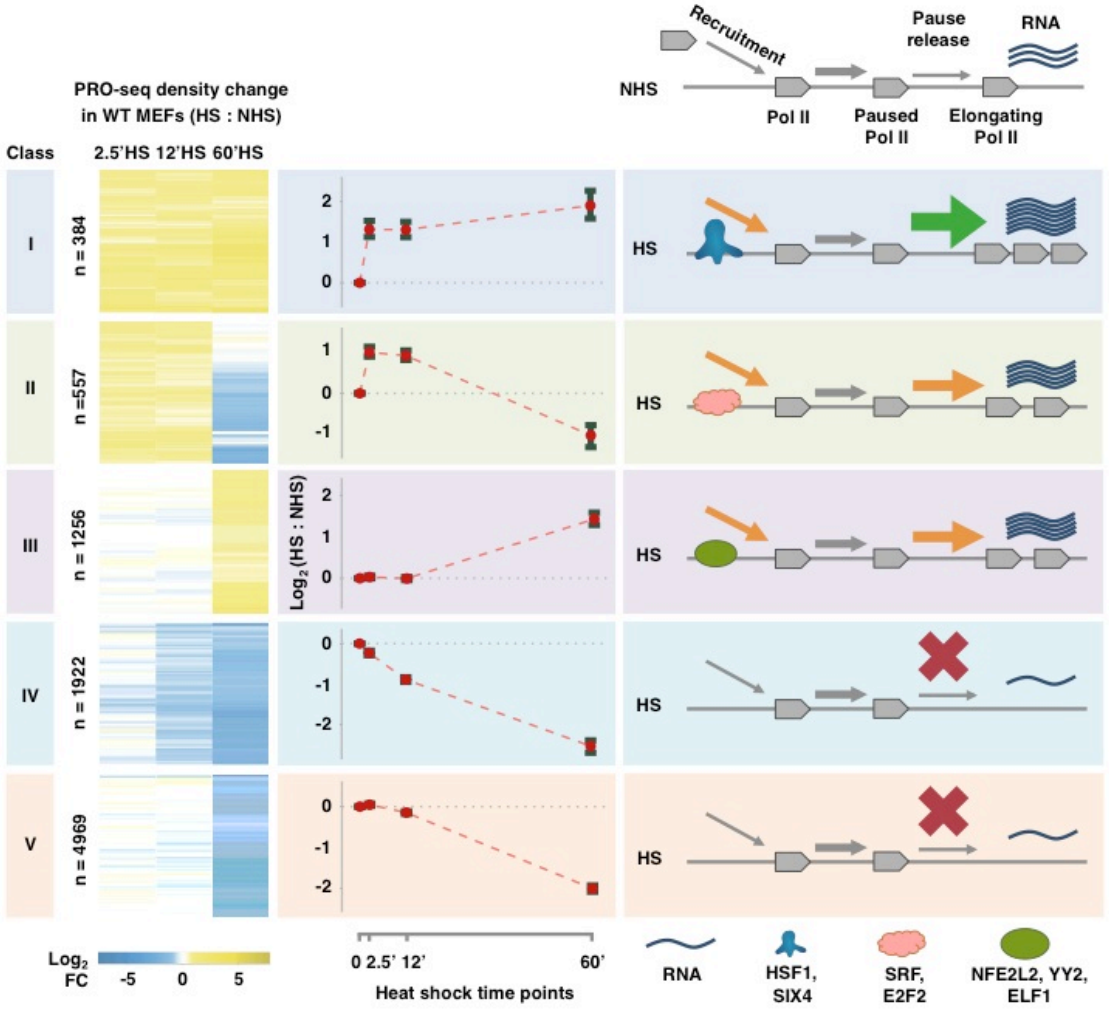


Figure 2.18. Kinetics and dynamics of transcription induction and repression during HSR

Kinetics, dynamics, and the proposed mechanism of the five major classes of HS-regulated genes in WT MEFs. Heatmap on the left panel shows the gene-body PRO-seq density change upon HS for all genes in each class. The mid panels show the average change in gene-body PRO-seq density - the red points represent the average $\log_2$ fold change for all genes and the green error bars represent 95% confidence interval. The right panels show the step(s) in transcription being regulated. Green arrow indicates positive regulation, the orange arrow represents likely but uncertain regulation, red cross indicates negative regulation, and thickness of the arrows indicates the extent of regulation. The most highly enriched TF motifs in different classes are shown at the bottom.

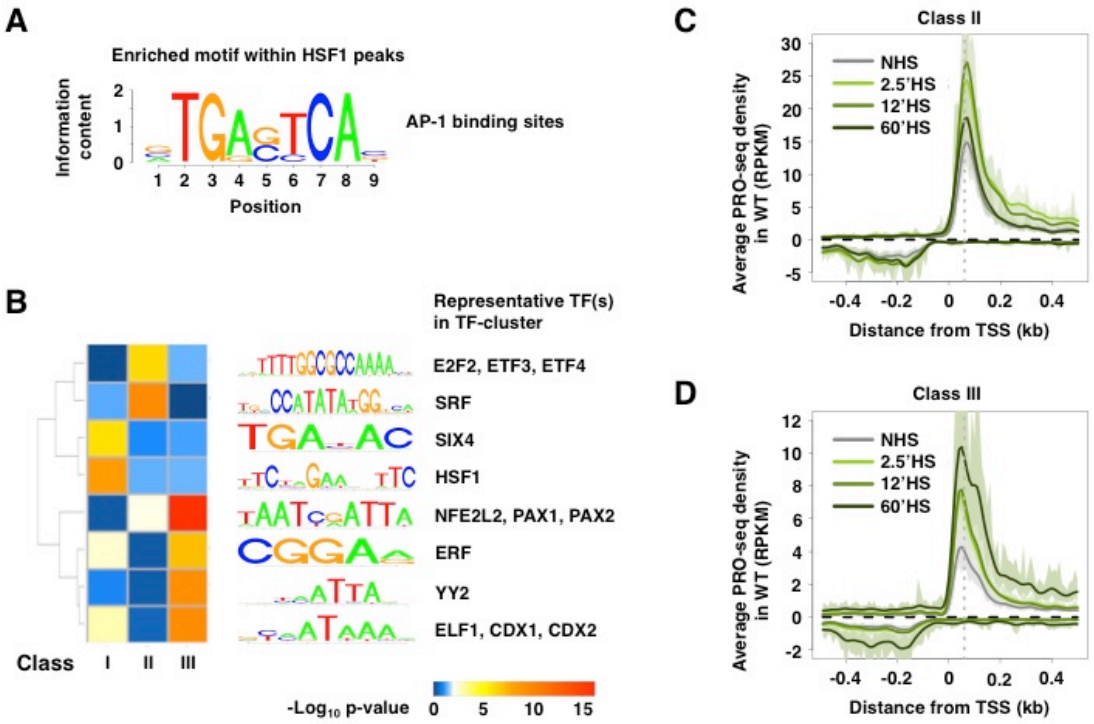


Figure 2.19. Kinetic classes have different TF motifs enriched in their promoters

(A) AP-1 TF motif is highly enriched within the HSF1 peaks.

(B) Significantly represented TF motifs out of 1200 different TF motifs scanned in the promoter region of the three kinetic classes of upregulated genes in WT MEFs. P-value of motif enrichment over 3rd order Markov model is shown in heatmap in negative Log_{10} scale (left) with sequence logo of most enriched motif (mid), and the TFs that bind to sequences similar to the motifs.

(C) PRO-seq density (with 95% confidence interval in light shades) around the TSS of transiently induced genes (class II) upon HS.

(D) PRO-seq density (with 95% confidence interval in light shades) around the TSS of late induced genes (class III) upon HS.

targets of TFs that are induced in the earlier phase of HSR. Promoters of genes in this class are enriched in antioxidant binding elements where Nuclear Factor, Erythroid 2-Like 2 (NFE2L2) is known to bind and induce oxidative stress response genes (Li and Kong, 2009). Unlike class II genes, PRO-seq density in the pause regions of these genes increases over the course of HS (Figure 2.19D), while increase in the gene-body only occurs at the 60'HS.

Class IV comprises genes that are downregulated immediately upon HS. These genes are characterized by progressive reduction in PRO-seq density over the course of HS due to the inhibition of Pol II release into productive elongation from the promoter-proximal paused state. This class is enriched for genes related to metabolism, cell cycle, and protein synthesis.

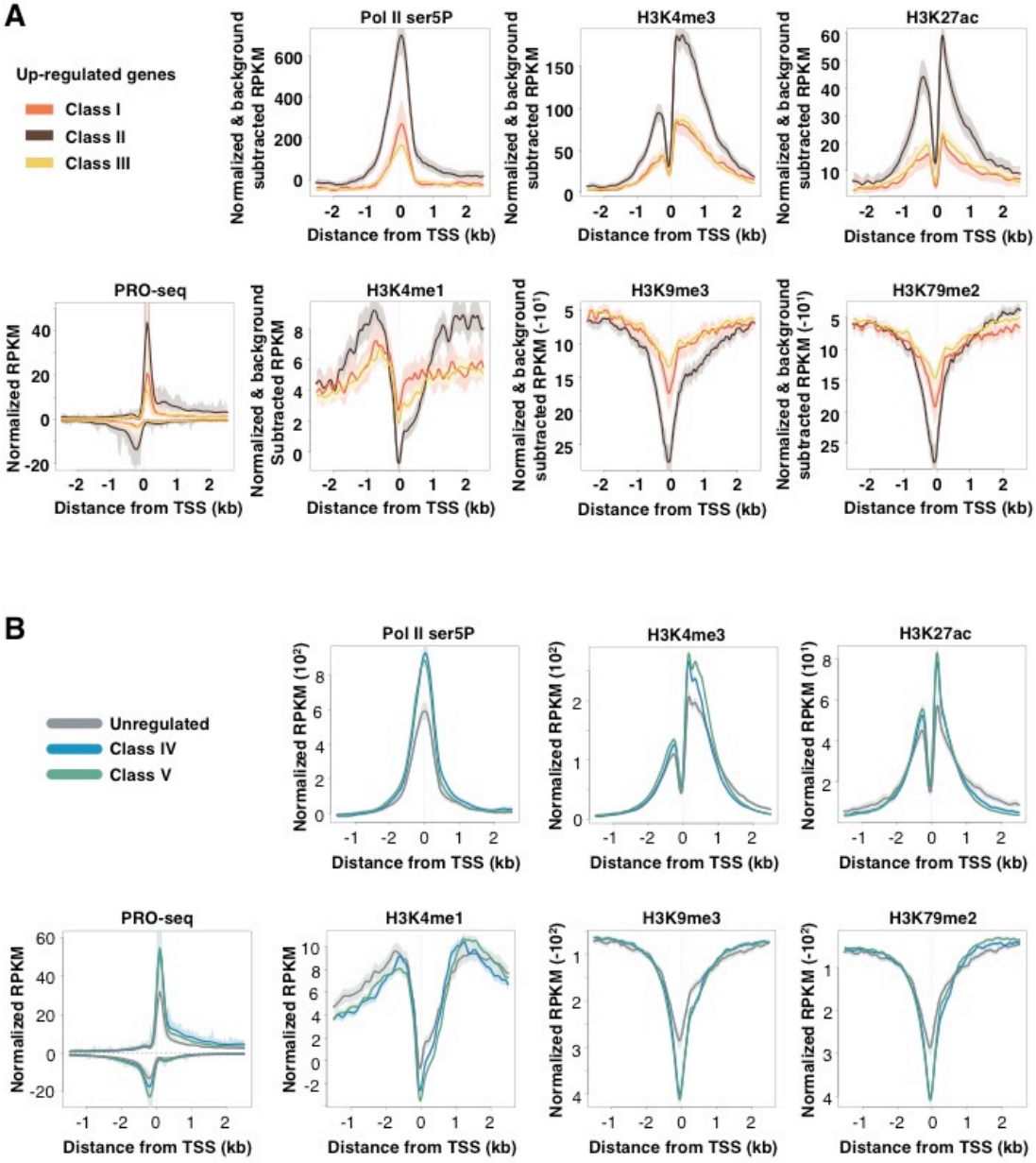
Class V consists genes that are unaffected during the early phase of the HSR but are downregulated in the late phase. Similar to the class IV, these genes are also regulated at the level of pause release. Interestingly, genes in this class are highly enriched for processes like splicing, mRNA processing, and nuclear transport.

To distinguish the characteristic features of five kinetic classes, we probed for all published and ENCODE-deposited genome-wide data in MEFs. We found that the kinetic classes are differentially enriched for several histone marks and Pol II prior to HS (Figures 2.20A and 2.20B). The class II genes are primed with higher levels of paused Pol II and higher permissive histone modifications compared to both class I and class III. In contrast, we did not

Figure 2.20. Kinetic classes have different chromatin modification enriched in their promoters

(A) Level of different histone marks and Ser5 Pol II (with 95% confidence interval in light shades) prior to HS around the TSS of three kinetic classes of transcriptionally-induced genes upon HS. PRO-seq density around the TSS for those genes is also shown.

(B) Level of different histone marks and Ser5 Pol II (with 95% confidence interval in light shades) prior to HS around the TSS of two kinetic classes of transcriptionally-repressed and unregulated genes upon HS. PRO-seq density around the TSS for those genes is also shown.


A
Up-regulated genes
Class I
Class II
Class III
Pol II ser5P
H3K4me3
H3K27ac
PRO-seq
H3K4me1
H3K9me3
H3K79me2
Normalized & background subtracted RPKM
Normalized RPKM
Distance from TSS (kb)
B
Unregulated
Class IV
Class V
Pol II ser5P
H3K4me3
H3K27ac
PRO-seq
H3K4me1
H3K9me3
H3K79me2
Normalized RPKM
Distance from TSS (kb)

find significant differences in histone modifications between the downregulated classes (class IV and class V). Overall, the differences in the kinetics and dynamics of regulation, in the associated TFs, Pol II, and histone marks, and in the specific biological processes targeted by each gene class indicate a highly regulated and carefully calibrated response mechanism that cells have evolved to cope with heat stress.

Discussion

Transcriptional regulation provides a major and primary level of defense from various proteotoxic stresses in cells from yeast to humans, yet the precise mechanism of global changes in gene expression during the HSR has remained unknown. Characterizing the complete network of regulated genes and understanding the molecular mechanisms involved in HSR are also critical for discovering novel therapeutic tools for cancer and neurodegenerative disorders as many players in the HSR are involved in these diseases. Here, we show that the HSR modulates an extensive fraction of the transcriptome, significantly upregulating 10% and downregulating 55% of the active genes at the level of transcription. HSR is extremely rapid, inducing changes in transcriptional patterns in as little as a minute and half, and different classes of genes exhibit diverse kinetics and dynamics of regulation. Mechanistically, we show that Pol II release from the promoter-proximal pause is a major regulatory step in both induction and repression of genes upon HS, and our

comprehensive analyses show that HSR is rapid, robust, and much more extensive than generally appreciated.

HSF1 is generally considered to be the master regulator of the HSR. In this study, while we demonstrate that HSF1 is critical for induction of over 250 genes, including the classical *Hsps*, we also show that HSF1 accounts for only a fraction of the transcriptionally-induced genes upon HS. Moreover, HSF1 has no detectable role in genome-wide repression. We demonstrate this by quantifying HS-induced changes in the density of transcribing Pol II complexes across the genome at high temporal and spatial (base pair) resolution using the extremely sensitive PRO-seq assay. We assessed the role of HSF1 both by performing a kinetic analysis of genome-wide changes in HSF1 binding upon HS using highly-specific antibodies in ChIP-seq, and by repeating the PRO-seq kinetic measurements in MEFs that have knock out of HSF1.

HSF1 is critically required during HS for induction of *Hsps* and over 200 other HSF1-dependent genes. HSF1 induces transcription in these genes by increasing the release of paused Pol II into productive elongation. These HSF1-dependent, upregulated genes also show concomitant increase in the level of paused Pol II; however this occurs in both WT and *Hsf1*$^{-/-}$ MEFs indicating the recruitment of Pol II to these genes during HS is mostly HSF1-independent. While this increased recruitment could be regulated by other HS-sensing TFs, we prefer the simple explanation that the increased availability of Pol II, which is caused by the global downregulation of transcription, drives this

Pol II loading onto the genes. Because global downregulation of transcription is also HSF1-independent, this increase availability of Pol II occurs in both WT and *Hsf1*$^{-/-}$ MEFs.

Interestingly, some genes bound by HSF1 in their promoters are not transcriptionally induced. This indicates that HSF1 that is bound near promoters requires a compatibility with other features of the promoter and its regulatory regions. This compatibility may require a promoter have both paused Pol II and features that recruit P-TEFb kinase, which is required generally for pause release in mammals (Jonkers et al., 2014; Price, 2000; Takahashi et al., 2011). In Drosophila polytene chromosomes, recruitment of the pause release factor P-TEFb to the heat shock loci was found to be dependent on HSF1 but this recruitment appeared not to be through a direct interaction with P-TEFb (Lis et al., 2000).). In this study, we observed an increase in P-TEFb level at the promoters of HS-induced genes and decrease in its level at the promoters of HS-repressed genes. We hypothesize that the nuclear P-TEFb level is redirected in HSF1-dependent manner to genes that are induced during HS. This active delivery of P-TEFb drains its free pool and results in reduced P-TEFb level at the promoters of non-critical genes resulting in their repression due to lack of pause release. This hypothesis can be tested by overexpression of P-TEFb. If the global redistribution of P-TEFb is driving the genome-wide transcription repression, we would expect rescue of transcription repression during HS due to higher availability of P-TEFb for non-

critical genes. This would be a critical step in understanding the precise interplay of HSF1 and other promoter elements and TFs in recruitment of P-TEFb.

In addition to promoters, HSF1 binds to many intragenic and intergenic sites. Intragenic HSF1 bound sites were previously proposed to repress transcription by creating a barrier in the gene-body to transcription. This particular hypothesis does not explain the bulk of downregulation observed in our study, as we find the intragenic HSF1 has very little influence on transcriptional downregulation on the bound gene. The intragenic-bound HSF1 bound genes that show transcriptional repression in WT, are also repressed in *Hsf1*$^{-/-}$ MEFs with comparable magnitude and kinetics. We further show that the density of Pol II upstream of the intragenic-bound HSF1 does not increase, indicating gene-bound HSF1 does not create appreciable impediment to elongating Pol II.

Intergenic-bound HSF1 is difficult to assign to a target gene without a high-resolution chromatin interaction map. Nevertheless, the divergent PRO-seq profile around the intergenic-bound HSF1 and the presence of HSF1-dependent genes without promoter bound HSF1 indicates that at least some of the intergenic-bound HSF1 functions as enhancer. The 150 HSF1-dependent genes that are not bound in their promoter region by HSF1 are potential targets of enhancer-bound HSF1. The transcription induction of these genes also occurs by an HSF1-dependent increase in pause release. This further

strengthens the role of HSF1 in HSR as a factor that increases the release of paused Pol II into productive elongation.

Theoretically, enhancer-bound HSF1 could not only induce but also repress transcription. We identify 232 genes in this study that have no detectable HSF1 binding but are induced at 60'HS in *Hsf1*$^{-/-}$ MEFs only. These genes could either be directly repressed by enhancer-bound HSF1, which is very unlikely given the absence of a documented role of HSF1 in repression, or indirectly repressed by HSF1 through HSF1-dependent TFs. While many of the genes induced in *Hsf1*$^{-/-}$ MEFs encode subunits of ATPase machinery, understanding the role of HSF1 in induction of these genes upon HS and the precise mechanism of their induction require study beyond the scope of this thesis. Nevertheless, our work redefines the role of HSF1 as the master regulator of HSR and demonstrates that the function of HSF1 is specialized in the robust activation of *Hsps* and over 200 other genes during HSR. This study also identifies over 900 genes that are induced upon HS in HSF1-independent manner and identifies additional TFs as potential activators of HSR.

One such TF that appears to play a large role in mediating the HSR is SRF. Arsenite-mediated stress is known to activate SRF via MAPK pathway, (Heidenreich et al., 1999); however, SRF's role in HSR was completely uncharacterized until now. The DNA-binding element of SRF is exceptionally enriched in promoters of transiently-induced genes during HSR, majority of which are cytoskeleton proteins. We hypothesize that the dramatic

cytoskeletal changes at the onset of HSR is the trigger that selectively induce genes in cytoskeleton family via the action of SRF. SRF activity depends on the MAL cofactor, which also binds to monomeric actin, and this actin binding sequesters MAL to the cytoplasm (Miralles et al., 2003; Vartiainen et al., 2007). In this model, the cytoskeletal polymerization caused by HS reduces the level of monomeric actin allowing MAL to form a complex with SRF and activate SRF-responsive genes. The transient induction of these genes could be mediated by the known feed-back loop where monomeric actin expression driven by SRF inhibits MAL association and activation of SRF (Sotiropoulos et al., 1999).

Transcriptional downregulation in response to HSR is one of the major regulatory adaptations proposed to reduce protein synthesis and misfolded protein aggregates (Jamrich et al., 1977; Lindquist, 1986). Our results indicate that the global downregulation is mediated by inhibition of promoter-proximal paused Pol II release into productive elongation and is independent of HSF1. However, the recruitment of Pol II to the downregulated genes is generally unaffected during HS causing an increase in paused Pol II, in contrast to a previous study in *Drosophila* cells that reported a reduction in paused Pol II levels upon HS, perhaps reflecting an evolutionary plasticity in the transcription steps that can be targeted (Teves and Henikoff, 2011). Identification of pause escape as a targeted regulatory step in transcription, provides a set of candidate factors associated with Pol II release from the

pause for future explorations of the mechanisms of repression during HS. By examining the mechanism of induction and repression, we demonstrate that the rapid HS-induced changes in PRO-seq density for both induced and repressed genes emanate from the promoter-associated pause region as a wave. Therefore, the increase and decrease in transcriptionally-engaged Pol II densities are not a consequence of thermally induced effects on Pol II elongation rates or processivity, but rather to changes on the rates of Pol II release from the pause.

Previous studies probing the genome-wide transcriptome regulated during HSR were unable to observe the primary transcriptional kinetics due to the requirement of longer HS treatment duration to accumulate or deplete polyadenylated mRNAs. Our approach of sequencing nascent RNAs before and after HS at multiple time points revealed that the transcription regulation is not a monolithic response. It comprises several classes of regulated genes with characteristic kinetics and dynamics. For instance, some genes are transcriptionally induced very early during HSR and they continue to be induced. Some are transiently induced and others are induced after a certain delay. Many in the latter class are likely the target of TFs induced in the early phase of HSR or late-binding HSF1. Identification of the precise step in transcription regulated in transient and late induction of genes requires further exploration. Similarly, downregulation also occurs in two distinct phases: early repressed genes encode proteins involved in cell cycle and metabolism and

late repressed genes encode proteins involved in mRNA processing. Post-transcriptional mRNA splicing is compromised after two hours of HS in MEFs (Shalgi et al., 2014), and our results indicate that a downregulation of splicing factors within the first hour of HS could be a contributing mechanism.

In summary, this study demonstrates the extensive nature of transcriptional regulation during HSR, where many hundreds of genes are activated and thousands repressed and with different classes of genes showing remarkably rapid and distinct kinetic profiles of regulation. This study also redefines the prevailing opinion that HSF1 is the master regulator of HSR: it demonstrates HSF1 as one major regulator specializing in robust induction of only a subset of genes regulated during HSR, and it identifies additional TFs as candidate regulators of HSR. The profiles of Pol II in wild type and *Hsf1*$^{-/-}$ MEFs also establishes that the release from promoter-proximal pause is the mechanistic step in transcription induction targeted by HSF1, and in global HSF1-independent repression of genes. Finally, the insights into the mechanism of HSF1 action and its primary gene targets will hopefully prove useful in the development of HSF1-based therapeutics for cancer and neurodegenerative disorders.

REFERENCES

Akerfelt, M., Morimoto, R.I., and Sistonen, L. (2010). Heat shock factors: integrators of cell stress, development and lifespan. Nat Rev Mol Cell Biol *11*, 545–555.

Anckar, J., and Sistonen, L. (2011). Regulation of HSF1 function in the heat stress response: implications in aging and disease. Annu. Rev. Biochem. *80*, 1089–1115.

Anders, S., Pyl, P.T., and Huber, W. (2015). HTSeq--a Python framework to work with high-throughput sequencing data. Bioinformatics *31*, 166–169.

Ardehali, M.B., and Lis, J.T. (2009). Tracking rates of transcription and splicing in vivo. Nat. Struct. Mol. Biol. *16*, 1123–1124.

Bailey, T., Krajewski, P., Ladunga, I., Lefebvre, C., Li, Q., Liu, T., Madrigal, P., Taslim, C., and Zhang, J. (2013). Practical guidelines for the comprehensive analysis of ChIP-seq data. PLoS Comput. Biol. *9*, e1003326.

Baird, N.A., Douglas, P.M., Simic, M.S., Grant, A.R., Moresco, J.J., Wolff, S.C., Yates, J.R., Manning, G., and Dillin, A. (2014). HSF-1-mediated cytoskeletal integrity determines thermotolerance and life span. Science *346*, 360–363.

Bell, J.L., Haak, A.J., Wade, S.M., Kirchhoff, P.D., Neubig, R.R., and Larsen, S.D. (2013). Optimization of novel nipecotic bis(amide) inhibitors of the Rho/MKL1/SRF transcriptional pathway as potential anti-metastasis agents. Bioorg. Med. Chem. Lett. *23*, 3826–3832.

Berendes, H.D. (1968). Factors involved in the expression of gene activity in polytene chromosomes. Chromosoma *24*, 418–437.

Bienz, M., and Pelham, H.R. (1986). Heat shock regulatory elements function as an inducible enhancer in the Xenopus hsp70 gene and when linked to a heterologous promoter. Cell *45*, 753–760.

Brown, J.B., Boley, N., Eisman, R., May, G.E., Stoiber, M.H., Duff, M.O., Booth, B.W., Wen, J., Park, S., Suzuki, A.M., et al. (2014). Diversity and dynamics of the Drosophila transcriptome. Nature *512*, 393–399.

Buckley, M.S., Kwak, H., Zipfel, W.R., and Lis, J.T. (2014). Kinetics of promoter Pol II on Hsp70 reveal stable pausing and key insights into its regulation. Genes & Development *28*, 14–19.

Chen, Y., Negre, N., Li, Q., Mieczkowska, J.O., Slattery, M., Liu, T., Zhang, Y., Kim, T.-K., He, H.H., Zieba, J., et al. (2012). Systematic evaluation of factors influencing ChIP-seq fidelity. Nature Publishing Group *9*, 609–614.

Core, L.J., Martins, A.L., Danko, C.G., Waters, C.T., Siepel, A., and Lis, J.T. (2014). Analysis of nascent RNA identifies a unified architecture of initiation regions at mammalian promoters and enhancers. Nat Genet *46*, 1311–1320.

Core, L.J., Waterfall, J.J., and Lis, J.T. (2008). Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. Science *322*, 1845–1848.

Dai, C., Whitesell, L., Rogers, A.B., and Lindquist, S. (2007). Heat shock factor 1 is a powerful multifaceted modifier of carcinogenesis. Cell *130*, 1005–1018.

Danko, C.G., Hah, N., Luo, X., Martins, A.L., Core, L., Lis, J.T., Siepel, A., and Kraus, W.L. (2013). Signaling pathways differentially affect RNA polymerase II initiation, pausing, and elongation rate in cells. Molecular Cell *50*, 212–222.

Danko, C.G., Hyland, S.L., Core, L.J., Martins, A.L., Waters, C.T., Lee, H.W., Cheung, V.G., Kraus, W.L., Lis, J.T., and Siepel, A. (2015). Identification of active transcriptional regulatory elements from GRO-seq data. Nature Publishing Group *12*, 433–438.

ENCODE Project Consortium (2012). An integrated encyclopedia of DNA elements in the human genome. Nature *489*, 57–74.

Esnault, C., Stewart, A., Gualdrini, F., East, P., Horswell, S., Matthews, N.,

and Treisman, R. (2014). Rho-actin signaling to the MRTF coactivators dominates the immediate transcriptional response to serum in fibroblasts. Genes & Development *28*, 943–958.

Fuda, N.J., Ardehali, M.B., and Lis, J.T. (2009). Defining mechanisms that regulate RNA polymerase II transcription in vivo. Nature *461*, 186–192.

Grant, C.E., Bailey, T.L., and Noble, W.S. (2011). FIMO: scanning for occurrences of a given motif. Bioinformatics *27*, 1017–1018.

Guertin, M.J., and Lis, J.T. (2010). Chromatin landscape dictates HSF binding to target DNA elements. PLoS Genet. *6*, e1001114.

Haak, A.J., Tsou, P.-S., Amin, M.A., Ruth, J.H., Campbell, P., Fox, D.A., Khanna, D., Larsen, S.D., and Neubig, R.R. (2014). Targeting the myofibroblast genetic switch: inhibitors of myocardin-related transcription factor/serum response factor-regulated gene transcription prevent fibrosis in a murine model of skin injury. J. Pharmacol. Exp. Ther. *349*, 480–486.

Hah, N., Danko, C.G., Core, L., Waterfall, J.J., Siepel, A., Lis, J.T., and Kraus, W.L. (2011). A rapid, extensive, and transient transcriptional response to estrogen signaling in breast cancer cells. Cell *145*, 622–634.

Heidenreich, O., Neininger, A., Schratt, G., Zinck, R., Cahill, M.A., Engel, K., Kotlyarov, A., Kraft, R., Kostka, S., Gaestel, M., et al. (1999). MAPKAP kinase 2 phosphorylates serum response factor in vitro and in vivo. J. Biol. Chem. *274*, 14434–14443.

Huang, D.W., Sherman, B.T., and Lempicki, R.A. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc *4*, 44–57.

Inouye, S., Katsuki, K., Izu, H., Fujimoto, M., Sugahara, K., Yamada, S.I., Shinkai, Y., Oka, Y., Katoh, Y., and Nakai, A. (2003). Activation of Heat Shock Genes Is Not Necessary for Protection by Heat Shock Transcription Factor 1 against Cell Death Due to a Single Exposure to High Temperatures. Mol. Cell. Biol. *23*, 5882–5895.

Jamrich, M., Greenleaf, A.L., and Bautz, E.K. (1977). Localization of RNA polymerase in polytene chromosomes of Drosophila melanogaster. Proc. Natl. Acad. Sci. U.S.a. *74*, 2079–2083.

Jolma, A., Yan, J., Whitington, T., Toivonen, J., Nitta, K.R., Rastas, P., Morgunova, E., Enge, M., Taipale, M., Wei, G., et al. (2013). DNA-binding specificities of human transcription factors. Cell *152*, 327–339.

Jonkers, I., and Lis, J.T. (2015). Getting up to speed with transcription elongation by RNA polymerase II. Nat Rev Mol Cell Biol *16*, 167–177.

Jonkers, I., Kwak, H., and Lis, J.T. (2014). Genome-wide dynamics of Pol II elongation and its interplay with promoter proximal pausing, chromatin, and exons. Elife *3*, e02407.

Kampinga, H.H., Hageman, J., Vos, M.J., Kubota, H., Tanguay, R.M., Bruford, E.A., Cheetham, M.E., Chen, B., and Hightower, L.E. (2009). Guidelines for the nomenclature of the human heat shock proteins. Cell Stress Chaperones *14*, 105–111.

Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., and Salzberg, S.L. (2013). TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol. *14*, R36.

Kim, T.-K., Hemberg, M., Gray, J.M., Costa, A.M., Bear, D.M., Wu, J., Harmin, D.A., Laptewicz, M., Barbara-Haley, K., Kuersten, S., et al. (2010). Widespread transcription at neuronal activity-regulated enhancers. Nature *465*, 182–187.

Kwak, H., Fuda, N.J., Core, L.J., and Lis, J.T. (2013). Precise maps of RNA polymerase reveal how promoters direct initiation and pausing. Science *339*, 950–953.

la Grange, de, P., Dutertre, M., Martin, N., and Auboeuf, D. (2005). FAST DB: a website resource for the study of the expression regulation of human gene products. Nucleic Acids Research *33*, 4276–4284.

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol.

Laszlo, A. (1992). The effects of hyperthermia on mammalian cell structure and function. Cell Prolif. *25*, 59–87.

Lecomte, S., Desmots, F., Le Masson, F., Le Goff, P., Michel, D., Christians, E.S., and Le Dr eacute an, Y. (2010). Roles of heat shock factor 1 and 2 in response to proteasome inhibition: consequence on p53 stability. Oncogene *29*, 4216–4224.

Lecomte, S., Reverdy, L., Le Quément, C., Le Masson, F., Amon, A., le Goff, P., Michel, D., Christians, E., and le Dréan, Y. (2013). Unraveling complex interplay between heat shock factor 1 and 2 splicing isoforms. PLoS ONE *8*, e56085.

Li, W., and Kong, A.-N. (2009). Molecular mechanisms of Nrf2-mediated antioxidant response. Mol. Carcinog. *48*, 91–104.

Lindquist, S. (1986). The heat-shock response. Annu. Rev. Biochem. *55*, 1151–1191.

Lindquist, S. (1987). Translational regulation in the heat-shock response of Drosophila cells. Translational Regulation of Gene Expression.

Lindquist, S., and Craig, E.A. (1988). The heat-shock proteins. Annu. Rev. Genet. *22*, 631–677.

Lis, J.T., Mason, P., Peng, J., Price, D.H., and Werner, J. (2000). P-TEFb kinase recruitment and function at heat shock loci. Genes & Development *14*, 792–803.

Locke, M., and Tanguay, R.M. (1996). Diminished heat shock response in the aged myocardium. Cell Stress Chaperones *1*, 251–260.

Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. *15*, 550.

Ma, X., Xu, L., Alberobello, A.T., Gavrilova, O., Bagattin, A., Skarulis, M., Liu, J., Finkel, T., and Mueller, E. (2015). Celastrol Protects against Obesity and Metabolic Dysfunction through Activation of a HSF1-PGC1α Transcriptional Axis. Cell Metab. *22*, 695–708.

Marcel, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet Journal *17*, 10–12.

Mathelier, A., Zhao, X., Zhang, A.W., Parcy, F., Worsley-Hunt, R., Arenillas, D.J., Buchman, S., Chen, C.-Y., Chou, A., Ienasescu, H., et al. (2014). JASPAR 2014: an extensively expanded and updated open-access database of transcription factor binding profiles. Nucleic Acids Research *42*, D142–D147.

McMillan, D.R., Christians, E., Forster, M., Xiao, X., Connell, P., Plumier, J.C., Zuo, X., Richardson, J., Morgan, S., and Benjamin, I.J. (2002). Heat Shock Transcription Factor 2 Is Not Essential for Embryonic Development, Fertility, or Adult Cognitive and Psychomotor Function in Mice. Mol. Cell. Biol. *22*, 8005–8014.

McMillan, D.R., Xiao, X., Shao, L., Graves, K., and Benjamin, I.J. (1998). Targeted disruption of heat shock transcription factor 1 abolishes thermotolerance and protection against heat-inducible apoptosis. J. Biol. Chem. *273*, 7523–7528.

Mendillo, M.L., Santagata, S., Koeva, M., Bell, G.W., Hu, R., Tamimi, R.M., Fraenkel, E., Ince, T.A., Whitesell, L., and Lindquist, S. (2012). HSF1 Drives a Transcriptional Program Distinct from Heat Shock to Support Highly Malignant Human Cancers. Cell *150*, 549–562.

Milleron, R.S., and Bratton, S.B. (2006). Heat shock induces apoptosis independently of any known initiator caspase-activating complex. J. Biol. Chem. *281*, 16991–17000.

Miralles, F., Posern, G., Zaromytidou, A.-I., and Treisman, R. (2003). Actin dynamics control SRF activity by regulation of its coactivator MAL. Cell *113*, 329–342.

Morimoto, R.I. (1998). Regulation of the heat shock transcriptional response: cross talk between a family of heat shock factors, molecular chaperones, and negative regulators. Genes & Development *12*, 3788–3796.

Morley, J.F., and Morimoto, R.I. (2004). Regulation of longevity in Caenorhabditis elegans by heat shock factor and molecular chaperones. Mol. Biol. Cell *15*, 657–664.

Murshid, A., Eguchi, T., and Calderwood, S.K. (2013). Stress proteins in aging and life span. Int J Hyperthermia *29*, 442–447.

Neef, D.W., Jaeger, A.M., and Thiele, D.J. (2011). Heat shock transcription factor 1 as a therapeutic target in neurodegenerative diseases. Nat Rev Drug Discov *10*, 930–944.

Neef, D.W., Turski, M.L., and Thiele, D.J. (2010). Modulation of heat shock transcription factor 1 as a therapeutic target for small molecule intervention in neurodegenerative disease. PLoS Biol. *8*, e1000291.

Neph, S., Vierstra, J., Stergachis, A.B., Reynolds, A.P., Haugen, E., Vernot, B., Thurman, R.E., John, S., Sandstrom, R., Johnson, A.K., et al. (2012). An expansive human regulatory lexicon encoded in transcription factor footprints. Nature *489*, 83–90.

Ni, Z., Saunders, A., Fuda, N.J., Yao, J., Suarez, J.-R., Webb, W.W., and Lis, J.T. (2008). P-TEFb is critical for the maturation of RNA polymerase II into productive elongation in vivo. Mol. Cell. Biol. *28*, 1161–1170.

O'Brien, T., and Lis, J.T. (1991). RNA polymerase II pauses at the 5' end of the transcriptionally induced Drosophila hsp70 gene. Mol. Cell. Biol. *11*, 5285–5290.

Page, T.J., Sikder, D., Yang, L., Pluta, L., Wolfinger, R.D., Kodadek, T., and Thomas, R.S. (2006). Genome-wide analysis of human HSF1 signaling reveals a transcriptional program linked to cellular adaptation and survival. Mol Biosyst *2*, 627–639.

Parker, C.S., and Topol, J. (1984). A Drosophila RNA polymerase II transcription factor binds to the regulatory site of an hsp 70 gene. Cell *37*, 273–283.

Perisic, O., Xiao, H., and Lis, J.T. (1989). Stable binding of Drosophila heat shock factor to head-to-head and tail-to-tail repeats of a conserved 5 bp recognition unit. Cell *59*, 797–806.

Peterlin, B.M., and Price, D.H. (2006). Controlling the elongation phase of transcription with P-TEFb. Molecular Cell *23*, 297–305.

Pickrell, J.K., Gaffney, D.J., Gilad, Y., and Pritchard, J.K. (2011). False positive peaks in ChIP-seq and other sequencing-based functional assays caused by unannotated high copy number regions. Bioinformatics *27*, 2144–2146.

Price, D.H. (2000). P-TEFb, a cyclin-dependent kinase controlling elongation by RNA polymerase II. Mol. Cell. Biol.

Rallu, M., Loones, M., Lallemand, Y., Morimoto, R., Morange, M., and Mezger, V. (1997). Function and regulation of heat shock factor 2 during mouse embryogenesis. Proc. Natl. Acad. Sci. U.S.a. *94*, 2392–2397.

Raychaudhuri, S., Loew, C., Körner, R., Pinkert, S., Theis, M., Hayer-Hartl, M., Buchholz, F., and Hartl, F.U. (2014). Interplay of Acetyltransferase EP300 and the Proteasome System in Regulating Heat Shock Transcription Factor 1. Cell *156*, 975–985.

Renner, D.B., Yamaguchi, Y., Wada, T., Handa, H., and Price, D.H. (2001). A highly purified RNA polymerase II elongation control system. J. Biol. Chem. *276*, 42601–42609.

Ritossa, F. (1962). A new puffing pattern induced by temperature shock and DNP in Drosophila. Experientia.

Salamanca, H.H., Antonyak, M.A., Cerione, R.A., Shi, H., and Lis, J.T. (2014). Inhibiting heat shock factor 1 in human cancer cells with a potent RNA aptamer. PLoS ONE *9*, e96330.

Sarge, K.D., Zimarino, V., Holm, K., Wu, C., and Morimoto, R.I. (1991). Cloning and characterization of two mouse heat shock factors with distinct inducible and constitutive DNA-binding ability. Genes & Development *5*, 1902–1911.

Schratt, G., Weinhold, B., Lundberg, A.S., Schuck, S., Berger, J., Schwarz, H., Weinberg, R.A., Rüther, U., and Nordheim, A. (2001). Serum response factor is required for immediate-early gene activation yet is dispensable for proliferation of embryonic stem cells. Mol. Cell. Biol. *21*, 2933–2943.

Shalgi, R., Hurt, J.A., Lindquist, S., and Burge, C.B. (2014). Widespread Inhibition of Posttranscriptional Splicing Shapes the Cellular Transcriptome following Heat Shock. Cell Rep *7*, 1362–1370.

Sheng, M., and Greenberg, M.E. (1990). The regulation and function of c-fos and other immediate early genes in the nervous system. Neuron *4*, 477–485.

Sorger, P.K., Lewis, M.J., and Pelham, H.R. (1987). Heat shock factor is regulated differently in yeast and HeLa cells. Nature *329*, 81–84.

Sotiropoulos, A., Gineitis, D., Copeland, J., and Treisman, R. (1999). Signal-regulated activation of serum response factor is mediated by changes in actin dynamics. Cell *98*, 159–169.

Storti, R.V., Scott, M.P., Rich, A., and Pardue, M.L. (1980). Translational control of protein synthesis in response to heat shock in D. melanogaster cells. Cell *22*, 825–834.

Takahashi, H., Parmely, T.J., Sato, S., Tomomori-Sato, C., Banks, C.A.S.,

Kong, S.E., Szutorisz, H., Swanson, S.K., Martin-Brown, S., Washburn, M.P., et al. (2011). Human mediator subunit MED26 functions as a docking site for transcription elongation factors. Cell *146*, 92–104.

Teves, S.S., and Henikoff, S. (2011). Heat shock reduces stalled RNA polymerase II and nucleosome turnover genome-wide. Genes & Development *25*, 2387–2397.

Theodorakis, N.G., and Morimoto, R.I. (1987). Posttranscriptional regulation of hsp70 expression in human cells: effects of heat shock, inhibition of protein synthesis, and adenovirus infection on translation and mRNA stability. Mol. Cell. Biol. *7*, 4357–4368.

Trinklein, N.D., Murray, J.I., Hartman, S.J., Botstein, D., and Myers, R.M. (2004). The role of heat shock transcription factor 1 in the genome-wide regulation of the mammalian heat shock response. Mol. Biol. Cell *15*, 1254–1261.

Vartiainen, M.K., Guettler, S., Larijani, B., and Treisman, R. (2007). Nuclear actin regulates dynamic subcellular localization and activity of the SRF cofactor MAL. Science *316*, 1749–1752.

Weirauch, M.T., Yang, A., Albu, M., Cote, A.G., Montenegro-Montero, A., Drewe, P., Najafabadi, H.S., Lambert, S.A., Mann, I., Cook, K., et al. (2014). Determination and inference of eukaryotic transcription factor sequence specificity. Cell *158*, 1431–1443.

Westwood, J.T., Clos, J., and Wu, C. (1991). Stress-induced oligomerization and chromosomal relocalization of heat-shock factor. Nature *353*, 822–827.

Whitesell, L., and Lindquist, S. (2009). Inhibiting the transcription factor HSF1 as an anticancer strategy. Expert Opin. Ther. Targets *13*, 469–478.

Wu, C. (1984). Activating protein factor binds in vitro to upstream control sequences in heat shock gene chromatin. Nature *311*, 81–84.

Yost, H.J., Petersen, R.B., and Lindquist, S. (1990). 16 Posttranscriptional Regulation of Heat Shock Protein Synthesis in Drosophila. Cold Spring Harbor Monograph Archive *19*, 379–409.

Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W., et al. (2008). Model-based analysis of ChIP-Seq (MACS). Genome Biol. *9*, R137.

Zhou, J., Wan, J., Gao, X., Zhang, X., Jaffrey, S.R., and Qian, S.-B. (2015). Dynamic m(6)A mRNA methylation directs translational control of heat shock response. Nature *526*, 591–594.

Zhou, Q., Li, T., and Price, D.H. (2012). RNA polymerase II elongation control. Annu. Rev. Biochem.

Zobeck, K.L., Buckley, M.S., Zipfel, W.R., and Lis, J.T. (2010). Recruitment timing and dynamics of transcription factors at the Hsp70 loci in living cells. Molecular Cell *40*, 965–975.

CHAPTER 3[b]

ROLE OF HSF2 IN MAMMALIAN HEAT SHOCK RESPONSE

Summary

Transcription is one of the major nodes of regulation of the heat shock response orchestrated by the heat shock factor (HSF). Mammals have multiple HSFs: HSF1-4, and HSF1 is the primary inducer of heat shock protein genes (*Hsps*) that are responsible for restoration of proteostasis during stress. However, stress specific function of HSF2, which shares sequence homology to HSF1 and binds to the same DNA sequence, is poorly understood. Here, we show that HSF2 has a role in transcription repression but does not regulate transcription induction during HSR. Although HSF2 binds to the promoters of some heat-inducible *Hsps*, the binding is dependent on HSF1, and the induction of these *Hsps* is not dependent on HSF2. Strikingly, the vast majority of HSF2-dependent down-regulated genes lack HSF2 binding, suggesting an indirect role of HSF2 in repression. Together, this study decouples HSF2 from rapid and robust induction of *Hsps* during mammalian HSR, and elucidates its broader repressive role, which is likely mediated through other factors.

[b] Partially adapted from Mahat, D. B., Salamanca, H. H., Duarte, F. M., Danko, C. G., & Lis, J. T. (2016). Mammalian Heat Shock Response and Mechanisms Underlying Its Genome-wide Transcriptional Regulation. *Molecular Cell*, *62*(1), 63–78. Reprinted with permission from Elsevier.

Introduction

One major environmental change that organisms have to endure during their lifespan is temperature fluctuation. Exposure to elevated temperature could alter protein homeostasis and eventually disrupt optimal cellular environment. In order to cope with heat and other proteotoxic agents, species of all complexity from bacteria to humans have evolved with a highly conserved defense mechanism known as the heat shock response (HSR). This protective response is orchestrated at transcriptional level by heat shock factor (HSF).

Model organisms such as bacteria, yeast, worms, and fruit flies have a single HSF and are capable of mounting a robust HSR. However, mammalian model species such as mouse, rats, and humans have multiple HSFs: HSF1, HSF2, HSF3, and HSF4 (Akerfelt et al., 2010). HSF1 and HSF2 are ubiquitously expressed and share sequence homology (Ahn et al., 2001) and structural similarity (Jaeger et al., 2016; Neudegger et al., 2016). They bind to identical DNA sequence known as heat shock element (HSE) as homo- or hetero-trimer (Manuel et al., 2002; Sandqvist et al., 2009). Moreover, *Hsf2* gene is not essential for survival as *Hsf2*$^{-/-}$ mouse exhibit normal growth and viability (McMillan et al., 2002).

Despite HSF2's striking similarity with HSF1 and its apparent dispensability in organism's growth and survival, HSF2 has overcome evolutionary scrutiny of this apparent redundancy, indicating its importance in

non-critical cellular functions. Indeed, HSF2 is required for optimal brain development and differentiation of reproductive organs (Kallio et al., 2002). HSF2 occupies hundreds of loci in mitotic cells (Vihervaara et al., 2013), binds in a cluster and regulates genes in male-specific region of the Y chromosome (Akerfelt et al., 2008). It also binds to the promoters of selected *Hsps* under normal temperature and modulates their expression (Wilkerson et al., 2007) indicating its specialization in non-stress-related functions. However, HSF2 has a limited role in stress response. HSF2 heterotrimerizes with HSF1 upon HS and binds to the *Hsp70* promoter (Ostling et al., 2007). It preferentially binds to satellite III repeats in nuclear stress granules (Alastalo et al., 2003). Nevertheless, the genome-wide role of HSF2 in transcription regulation during HSR remains to be fully elucidated.

Understanding the contribution of HSF2, either by coordinating with HSF1 or as a compensatory factor of HSF1, in regulation of HSF1-dependnet genes during HS would underscore the evolutionary relevance of multiple HSFs in vertebrates. Similarly, if HSF2 has evolved to regulate a different set of genes than HSF1, identifying those genes and understanding the mechanism of their regulation would reveal the evolutionary rationale in diversification of heat shock transcription factor. More importantly, as modulation of HSF1's level and activity are being considered in treatment of neurological disorders and cancers (Neef et al., 2011; Whitesell and Lindquist, 2009), knowing HSF2-regulated genes, its mode of action, and its interaction

with HSF1 is critical in making informed decision on HSF1 inhibition or activation.

In this study, we examined the transcriptional roles of HSF2 in stress response with and without HSF1. We used precision nuclear run-on sequencing assay (PRO-seq) (Kwak et al., 2013) to measure nascent transcription prior to and at various time points during heat shock (HS) in *Hsf2*$^{-/-}$ and *Hsf1&2*$^{-/-}$ mouse embryonic fibroblasts (MEFs). To identified the HSF2-regulated genes and to decipher HSF2's role in global transcriptional changes during HS, we assessed the genome-wide transcriptional changes upon HS in these cell types and compared to our previously published study in WT and *Hsf1*$^{-/-}$ MEFs (Mahat et al., 2016). We carried out a time course of HSF2 chromatin immunoprecipitation and sequencing (ChIP-seq) in WT and *Hsf1*$^{-/-}$ MEFs to distinguish the primary targets of HSF2. We also dissected the role of HSF1 and HSF2 in regulation of *Hsps* and elucidated HSF2's dependence on HSF1 for DNA binding during stress. Although the precise mechanism of HSF2-mediated global repression remains to be resolved, the findings presented here reveal features of the mechanism and functions of HSF2 in regulation of transcription during HSR.

Materials and methods

Hsf2$^{-/-}$ *cell line*

Mouse embryonic fibroblasts were generated from *Hsf2*$^{-/-}$ mouse (McMillan et al., 2002) and were kindly gifted to us by Yves Le Dréan, Universite de Rennes 1, Rennes, France.

Data on wild type, Hsf1$^{-/-}$*, and Hsf1&2*$^{-/-}$ *MEFs*

PRO-seq and ChIP-seq data on wild type, *Hsf1*$^{-/-}$, and *Hsf1&2*$^{-/-}$ MEFs are taken from our previously published study (Mahat et al., 2016).

Cell culture, heat shock, and nuclei isolation

Cell culture, heat shock, and nuclei isolation were performed as described previously (Mahat et al., 2016). In short, MEFs were grown in Dulbecco's Modified Eagle Medium supplemented with 10% heat inactivated fetal bovine serum (v/v) and 1% Penicillin Streptomycin (v/v) at 37^0C with 5% CO_2 and 90% humidity.

Instantaneous HS was performed using ~80% confluent cells by adding pre-heated (42^0C) conditioned media collected from identically growing cells and by incubating the cell plates at 42^0C for the desired time. After the desired duration of HS, NHS cells grown at 37^0C and the heat shocked cells were harvested identically.

Nuclei isolation was done by douncing cells 25 times in cell membrane lysis buffer. The isolated nuclei were washed twice in lysis buffer and resuspended in storage buffer.

Nuclear run-on and PRO-seq library preparation

Nuclear run-on experiments were carried out by mixing 10×10^6 nuclei in 100μl of storage buffer with 100μl of 2x nuclear run-on buffer. The mix was incubated at 37^0C for three minutes. RNAs were isolated and base-hydrolyzed with 200 nM final concentration of NaOH to an average size between 100-150 nucleotides. Biotinylated nascent RNAs were isolated with magnetic beads coated with streptavidin followed by 3' adapter ligation. After another round of biotin-streptavidin affinity purification, the mRNA cap was removed and the 5' adapter was ligated. After the third biotin-streptavidin affinity purification, cDNA was generated by reverse transcription and PRO-seq libraries were prepared for sequencing using Illumina TruSeq small-RNA adaptors.

Mapping, normalization, and differential expression of PRO-seq libraries

PRO-seq reads were mapped against mm10 genome. Libraries were normalized using PRO-seq reads in 3' end of long genes. Differential expression of PRO-seq libraries were performed using DESeq2. These methods are described in detail previously (Mahat et al., 2016).

HSF2 ChIP-seq library preparation and peak calling

ChIP was performed as described previously with some modifications (Mahat et al., 2016). The HSF2 antibody was a generous gift of Lea Sistonen at University of Turku, Finland. HSF2 peaks were called against library

prepared with non-specific IgG using MACS2 (Zhang et al., 2008). In order to minimize the false positive peaks, we only considered HSF1 peaks that were identified with both antibodies at a threshold of $1e^{-05}$.

Results

Lack of HSF2 mitigates global downregulation during HSR

To decipher the role of HSF2 in HSR and to test whether HSF2 is compensating for the lack of HSF1 in *Hsf1*$^{-/-}$ MEFs, we performed PRO-seq in nuclei isolated from *Hsf2*$^{-/-}$ and *Hsf1&2*$^{-/-}$ MEFs at 0', 2.5', 12', and 60' after HS (Figure 3.1A and Table 3.1). *Hsf2*$^{-/-}$ MEFs were isolated from *Hsf2*$^{-/-}$ mouse that was generated by deletion of a 3 kb region in the 5' end of the gene including exon 1 (McMillan et al., 2002), and *Hsf1&2*$^{-/-}$ MEFs were isolated from mouse generated by crossing *Hsf1*$^{-/-}$ and *Hsf2*$^{-/-}$ mouse (Lecomte et al., 2010; 2013). PRO-seq profiles confirm targeted genomic deletion in *Hsf2* and *Hsf1* & *Hsf2* genes to make *Hsf2*$^{-/-}$ and *Hsf1&2*$^{-/-}$ mice respectively (Figure 3.1B). Two PRO-seq libraries were made for each cell type and condition, and the biological replicates correlated well (Figure 3.1C and Table 3.2).

We then quantified PRO-seq density in all genes and measured genome-wide change in transcription during HS. Similar to WT and *Hsf1*$^{-/-}$ MEFs, few genes are regulated as early as 2.5 min of HS in both *Hsf2*$^{-/-}$ and *Hsf1&2*$^{-/-}$ MEFs and the number of transcriptionally changed genes increases with HS time (Figure 3.2A). Although the overall profile of transcriptional

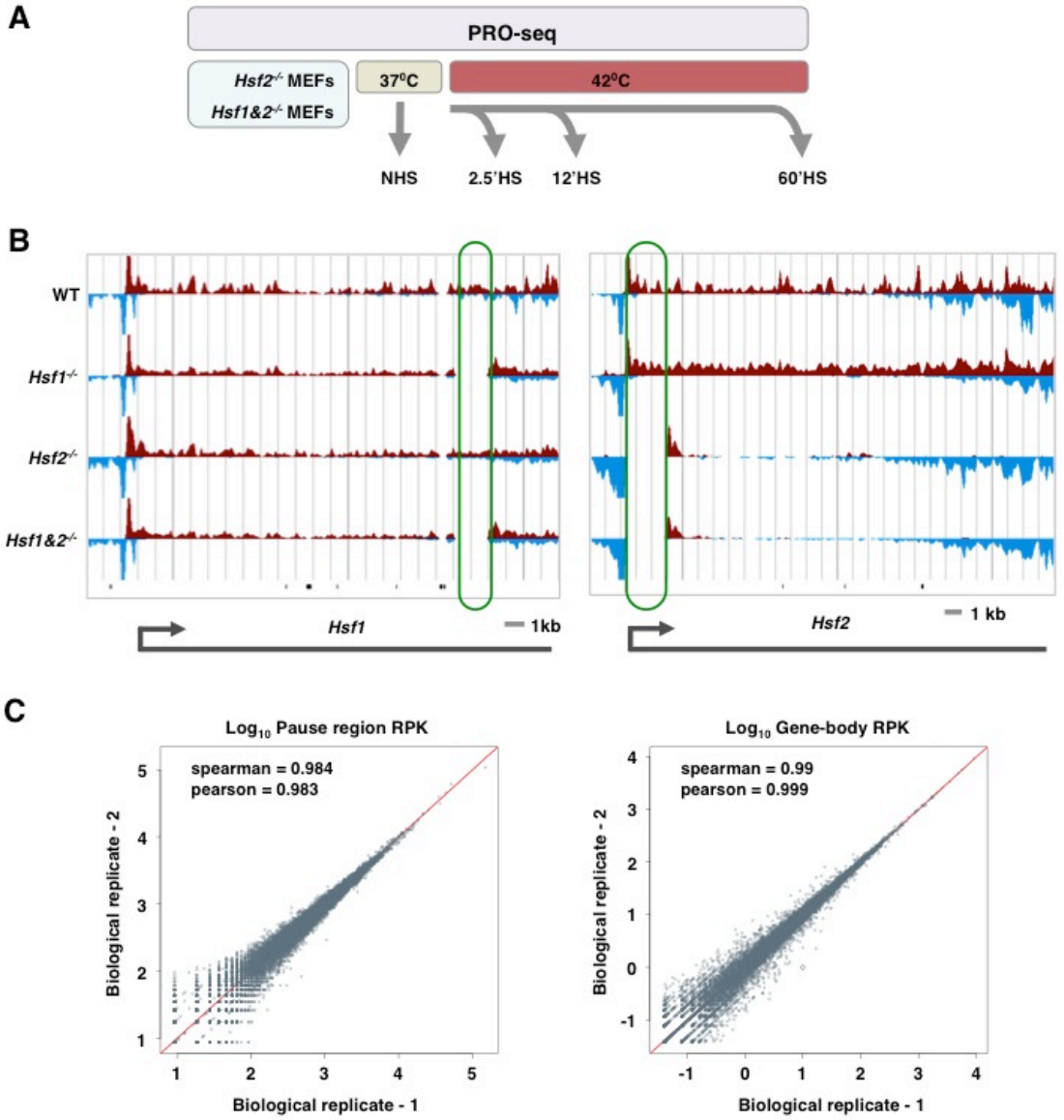


Figure 3.1. PRO-seq in *Hsf2*$^{-/-}$ and *Hsf1&2*$^{-/-}$ MEFs
(A) Experimental set-up, PRO-seq assay was performed in nuclei isolated from *Hsf2*$^{-/-}$ and *Hsf1&2*$^{-/-}$ MEFs at NHS, 2.5'HS, 12'HS, and 60'HS.
(B) Genome browser screenshot of *Hsf1* and *Hsf2* gene. The 2 kb region devoid of PRO-seq reads (green box) in *Hsf1* gene represents half of the DNA binding domain and trimerization domain excised to create *Hsf1*$^{-/-}$ MEFs. Similarly, the 3 kb region including the 1st exon of *Hsf2* gene devoid of PRO-seq reads (green box), was excised to create *Hsf2*$^{-/-}$ MEFs.
(C) A representative correlation plot between the biological replicates of PRO-seq libraries in the pause region (left) and the gene-body region. Complete table of Pearson and Spearman correlation values are listed in Table 3.2.

Table 3.1. Sequencing depth & alignment statistics of PRO-seq libraries.

Library	Sequenced	Unaligned (%)	Non-unique (%)	Aligned to mm10
***Hsf2*$^{-/-}$_NHS_BR1**	32,689,692	16.81	29.97	13,572,909
***Hsf2*$^{-/-}$_NHS_BR2**	34,713,509	20.27	26.50	15,449,435
***Hsf2*$^{-/-}$_2.5'HS_BR1**	34,068,909	18.58	30.08	13,279,281
***Hsf2*$^{-/-}$_2.5'HS_BR2**	34,866,606	17.73	32.03	12,460,068
***Hsf2*$^{-/-}$_12'HS_BR1**	36,229,989	16.97	30.67	14,824,815
***Hsf2*$^{-/-}$_12'HS_BR2**	35,287,736	17.82	30.41	14,322,306
***Hsf2*$^{-/-}$_60'HS_BR1**	35,282,523	12.95	30.79	15,522,593
***Hsf2*$^{-/-}$_60'HS_BR2**	36,252,164	13.17	30.41	16,167,571
***Hsf1&2*$^{-/-}$_NHS_BR1**	51,583,525	11.06	16.46	20,802,253
***Hsf1&2*$^{-/-}$_NHS_BR2**	49,804,354	11.67	12.17	26,097,930
***Hsf1&2*$^{-/-}$_2.5'HS_BR1**	93,139,446	3.57	23.88	14,574,252
***Hsf1&2*$^{-/-}$_2.5'HS_BR2**	76,230,283	3.49	30.35	15,304,152
***Hsf1&2*$^{-/-}$_12'HS_BR1**	61,272,461	9.43	15.40	24,686,166
***Hsf1&2*$^{-/-}$_12'HS_BR2**	89,042,556	9.71	14.22	36,080,133
***Hsf1&2*$^{-/-}$_60'HS_BR1**	91,113,987	9.18	17.64	39,617,211
***Hsf1&2*$^{-/-}$_60'HS_BR2**	57,704,222	7.51	20.76	22,096,193

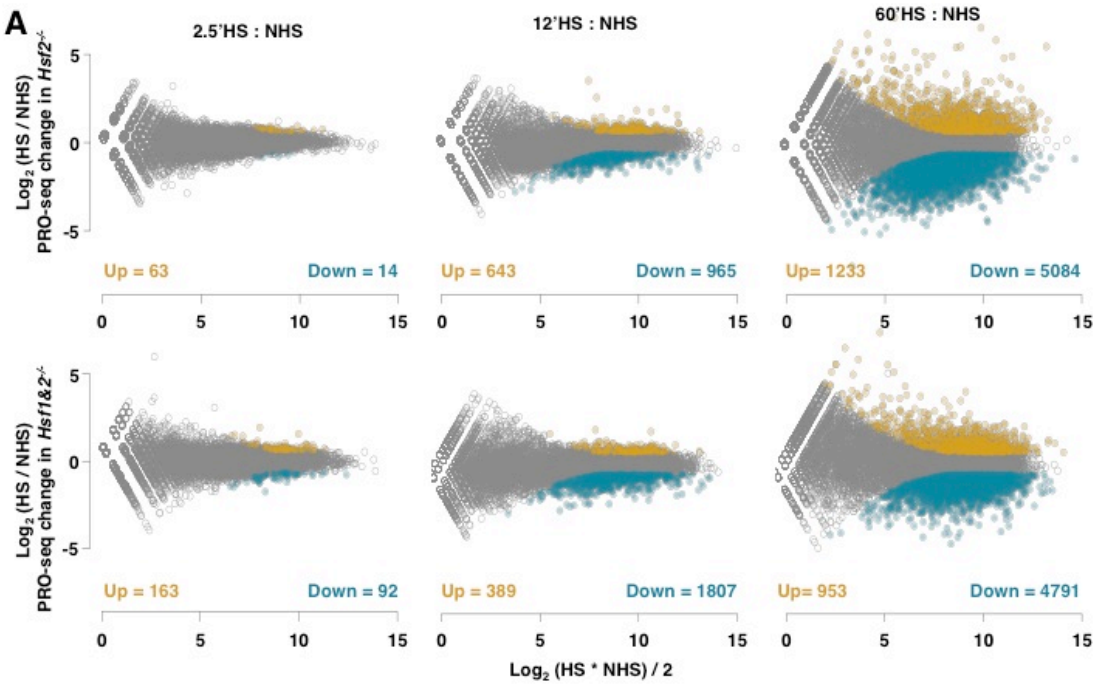
Table 3.2. Correlation between biological replicates of PRO-seq libraries.

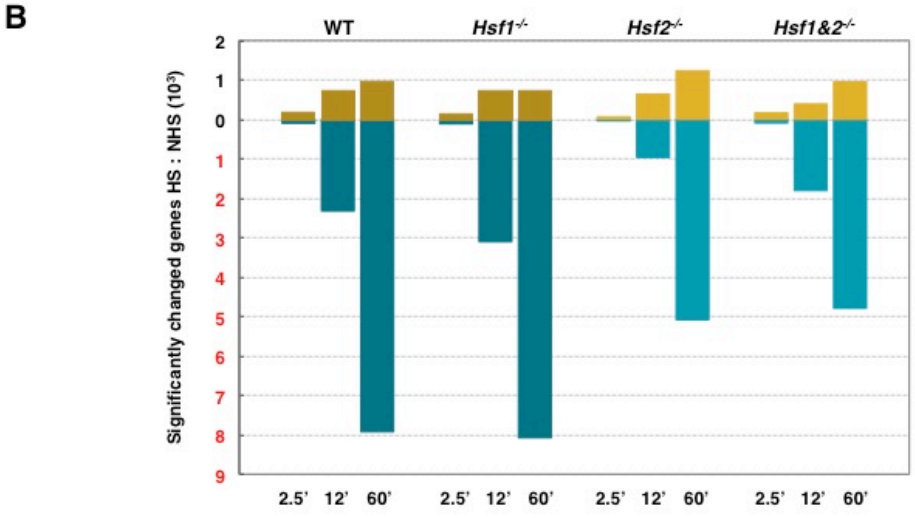
Library	Pearson correlation		Spearman correlation	
	Promoter	Gene body	Promoter	Gene body
***Hsf2*$^{-/-}$_NHS**	0.99	0.99	0.98	0.97
***Hsf2*$^{-/-}$_2.5'HS**	0.99	0.99	0.98	0.98
***Hsf2*$^{-/-}$_12'HS**	0.98	0.99	0.98	0.97
***Hsf2*$^{-/-}$_60'HS**	0.99	0.99	0.98	0.98
***Hsf1&2*$^{-/-}$_NHS**	0.99	0.99	0.98	0.97
***Hsf1&2*$^{-/-}$_2.5'HS**	0.99	0.99	0.98	0.98
***Hsf1&2*$^{-/-}$_12'HS**	0.98	0.99	0.98	0.97
***Hsf1&2*$^{-/-}$_60'HS**	0.99	0.99	0.98	0.98

Figure 3.2. More upregulated and less downregulated genes during HSR in absence of HSF2

(A) 'Minus-average' (MA) plots represent PRO-seq density change in the gene-body of all genes (n=23460) between NHS and 2.5'HS (left panels), 12'HS (mid panels), and 60'HS (right panels) in *Hsf2*$^{-/-}$ MEFs (top panels) and *Hsf1&2*$^{-/-}$ MEFs (bottom panels). Significantly upregulated genes (p-value < 0.001 in DESeq2) are shown in gold and significantly downregulated genes are shown in blue.

(B) Number of significantly changed genes upon HS in *Hsf2*$^{-/-}$ and *Hsf1&2*$^{-/-}$ MEFs at NHS, 2.5'HS, 12'HS, and 60'HS. Number of significantly changed genes upon HS in WT and *Hsf1*$^{-/-}$ MEFs are shown as well. Upregulated genes are shown in gold and downregulated genes are shown in blue.


A
2.5'HS : NHS
12'HS : NHS
60'HS : NHS
Log2 (HS / NHS) PRO-seq change in Hsf2-/-
Up = 63
Down = 14
Up = 643
Down = 965
Up= 1233
Down = 5084
Log2 (HS / NHS) PRO-seq change in Hsf1&2-/-
Up = 163
Down = 92
Up = 389
Down = 1807
Up= 953
Down = 4791
Log2 (HS * NHS) / 2

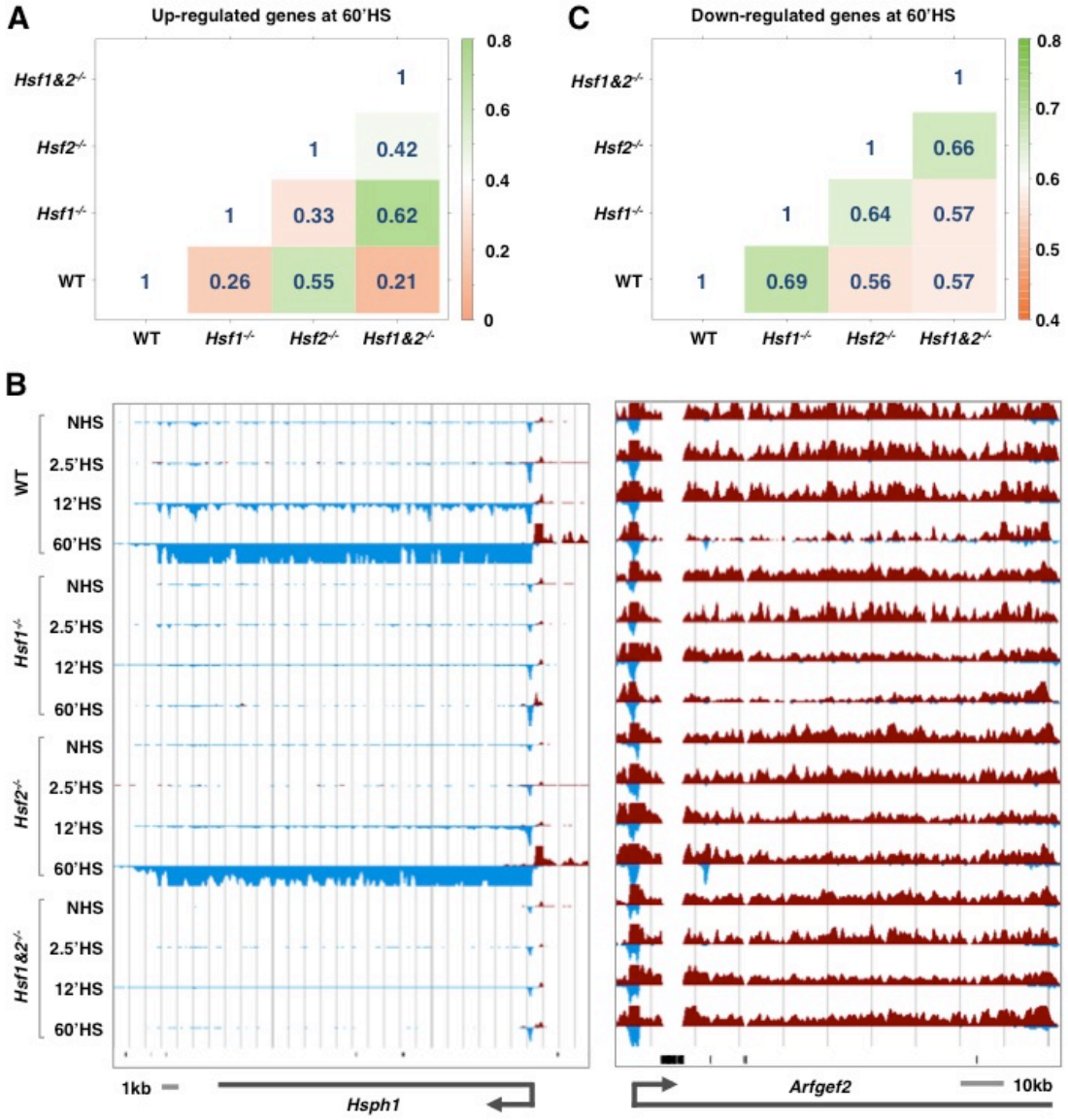

B
WT
Hsf1-/-
Hsf2-/-
Hsf1&2-/-
Significantly changed genes HS : NHS (10^3)
2.5' 12' 60'

change in *Hsf2*$^{-/-}$ and *Hsf1&2*$^{-/-}$ MEFs resemble the WT and *Hsf1*$^{-/-}$ MEFs, a closer inspection of actual number of changed genes at 60 min of HS, a duration for full-fledged HSR, and comparison with WT and *Hsf1*$^{-/-}$ MEFs reveal a higher number of genes were upregulated and fewer genes were downregulated in absence of HSF2 (Figure 3.2B). Similarly, comparison of *Hsf1&2*$^{-/-}$ with WT and *Hsf1*$^{-/-}$ MEFs also indicate that the decrease in number of upregulated genes in *Hsf1*$^{-/-}$ is compensated by further loss of HSF2 in *Hsf1&2*$^{-/-}$ MEFs, whereas HSF2's absence results in fewer downregulated genes (Figure 3.2B). The number of upregulated genes in *Hsf2*$^{-/-}$ is also higher than in *Hsf1&2*$^{-/-}$ MEFs; however, the pool of downeregulated genes is relatively unchanged (Figure 3.2B). Together, these observations suggest that the HSF1 is a transcription inducer, whereas lack of HSF2 mitigates global downregulation during HSR.

HSF2 plays a role in transcription repression during HS

To further investigate the opposing regulation of HSF1 and HSF2 during HSR, we correlated the HS-regulated genes, upregulated and downregulated separately, between all cell types. The upregulated genes have higher correlation between WT and *Hsf2*$^{-/-}$ and between *Hsf1*$^{-/-}$ and *Hsf1&2*$^{-/-}$ compared to other pairwise correlation, indicating HSF1's presence or absence respectively as a major determinant in the number and the extent of upregulated genes (Figure 3.3A). *Hsph1* gene is an example of HSF1-

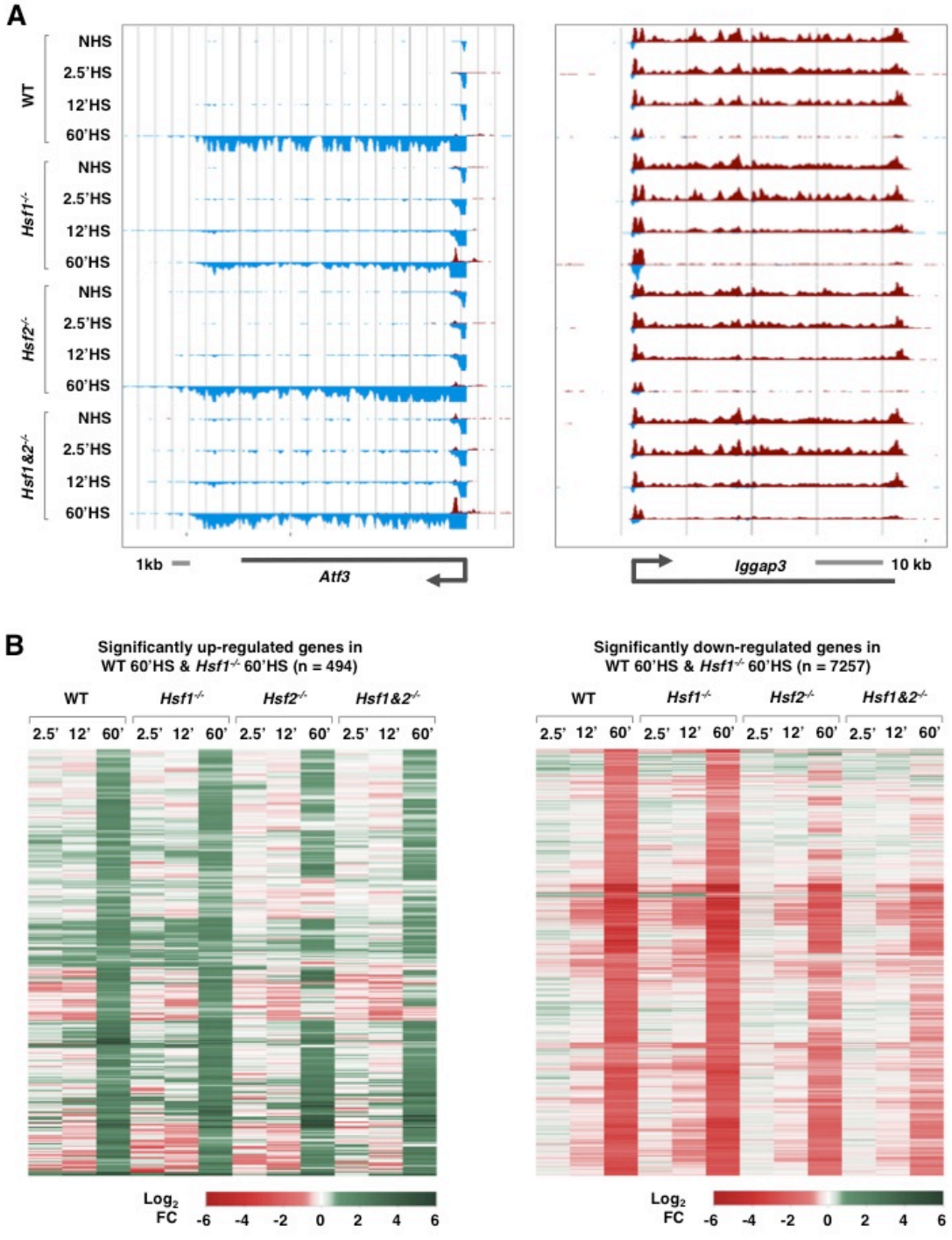
Figure 3.3. Divergent roles of HSF1 and HSF2 during HSR
(A) Correlation matrix showing the pairwise correlation between significantly upregulated genes in four cell types: WT, *Hsf1*$^{-/-}$, *Hsf2*$^{-/-}$, and *Hsf1&2*$^{-/-}$ MEFs. Correlation values are represented as heatmap and are also numerically stated.
(B) Screenshots of *Hsph1* gene representing HSF1-dependent induction (left panel) and *Arfgef2* gene representing HSF2-dependent repression (right panel). PRO-seq density in sense and antisense direction is shown in red and blue respectively. Small vertical black bars at the bottom of PRO-seq tracks represent the genomic regions that do not map uniquely at 36bp resolution.
(C) Correlation matrix showing the pairwise correlation between significantly downregulated genes in four cell types: WT, *Hsf1*$^{-/-}$, *Hsf2*$^{-/-}$, and *Hsf1&2*$^{-/-}$ MEFs. Correlation values are represented as heatmap and are also numerically stated.


A
Up-regulated genes at 60'HS
Hsf1&2-/-
Hsf2-/-
Hsf1-/-
WT
1
0.42
0.33
0.62
0.26
0.55
0.21
0.8
0.6
0.4
0.2
0
C
Down-regulated genes at 60'HS
0.66
0.64
0.57
0.69
0.56
0.57
0.7
0.5
0.4
B
NHS
2.5'HS
12'HS
60'HS
1kb
Hsph1
Arfgef2
10kb

dependent induction, which is induced in WT and *Hsf2*$^{-/-}$ but not in *Hsf1*$^{-/-}$ and *Hsf1&2*$^{-/-}$ MEFs (Figure 3.3B, left panel). In contrast, the downregulated genes have higher correlation between WT and *Hsf1*$^{-/-}$ and between *Hsf2*$^{-/-}$ and *Hsf1&2*$^{-/-}$, compared to other pairwise correlation, indicating HSF2's presence or absence respectively in this case as a key element in downregulation (Figure 3.3C). *Arfgef2* gene is an example of HSF2-mediated repression, which is downregulated in WT and *Hsf1*$^{-/-}$ but not in *Hsf2*$^{-/-}$ and *Hsf1&2*$^{-/-}$ MEFs (Figure 3.3B, right panel). These analyses indicate that HSF1 induces transcription during HSR but has no role in repression, which is also demonstrated by our earlier study of HSR in WT and *Hsf1*$^{-/-}$ MEFs (Mahat et al., 2016), whereas HSF2 plays a role in transcription repression and has little to no effect in induction.

Because HSF2 is present in both WT and *Hsf1*$^{-/-}$ MEFs, we examined whether similarly regulated genes in WT and *Hsf1*$^{-/-}$ MEFs (494 significantly upregulated and 7257 significantly downregulated genes) were in fact regulated by HSF2. We measured transcriptional changes on those genes in *Hsf2*$^{-/-}$ and *Hsf1&2*$^{-/-}$ and compared alongside WT and *Hsf1*$^{-/-}$ MEFs. Many genes similarly regulated in WT and *Hsf1*$^{-/-}$ show comparable regulation in *Hsf2*$^{-/-}$ and *Hsf1&2*$^{-/-}$ MEFs as shown by the two examples (Figure 3.4A). All upregulated genes in WT and *Hsf1*$^{-/-}$ are upregulated to the same extent in *Hsf2*$^{-/-}$ and *Hsf1&2*$^{-/-}$ (Figure 3.4B, left panel) indicating that they were not a result of HSF2-mediated compensation. However, the number of significantly

Figure 3.4. HSF2 likely plays a role in transcription repression during HS
(A) Screenshots of *Atf3* gene representing similarly induced genes in all cell types (left panel) and *Iggap3* gene representing similarly repressed genes in all cell types (right panel). PRO-seq density in sense and antisense direction is shown in red and blue respectively. Small vertical black bars at the bottom of PRO-seq tracks represent the genomic regions that do not map uniquely at 36bp resolution.
(B) Heatmap showing the transcriptional status in *Hsf2*$^{-/-}$ and *Hsf1&2*$^{-/-}$ MEFs upon HS for significantly upregulated genes (left) and significantly downregulated genes (right) in WT and *Hsf1*$^{-/-}$ MEFs.


A
WT
NHS
2.5'HS
12'HS
60'HS
Hsf1-/-
Hsf2-/-
Hsf1&2-/-
1kb
Atf3
Iggap3
10 kb
B
Significantly up-regulated genes in
WT 60'HS & Hsf1-/- 60'HS (n = 494)
Significantly down-regulated genes in
WT 60'HS & Hsf1-/- 60'HS (n = 7257)
2.5' 12' 60'
Log2 FC
-6 -4 -2 0 2 4 6

downregulated genes is lower in *Hsf2*$^{-/-}$ and *Hsf1&2*$^{-/-}$ compared to the WT and *Hsf1*$^{-/-}$ MEFs and the extent of downregulation is also relatively low (Figure 3.4B, right panel). This observation that some genes dowregulated in WT and *Hsf1*$^{-/-}$ are not downregulated in *Hsf2*$^{-/-}$ and *Hsf1&2*$^{-/-}$ strengthens our hypothesis that HSF2 mediates repression of at least some genes in WT and *Hsf1*$^{-/-}$ cells.

HSF2-dependent HS-regulated genes are not bound by HSF2

The HSF2-dependent regulation could either be a result of direct HSF2 binding to the promoters of regulated genes or a secondary effect mediated through other factors. To discern these possibilities, we probed genome-wide HSF2 binding during HS in WT and *Hsf1*$^{-/-}$ MEFs using ChIP-seq (Figure 3.5A and Table 3.3). The HSF2 antibody used in this ChIP-seq study is specific to mammalian HSF2 (Vihervaara et al., 2013), and as a control, we used non-specific IgG. We found that HSF2 binds to the promoters of heat-inducible *Hsps* such as *Hsph1* gene (Figure 3.5B), similar to HSF1 (Figure 3.5C) (Mahat et al., 2016). Interestingly, HSF2 binding at the promoter of *Hsph1* gene is severely limited in *Hsf1*$^{-/-}$ cells (Figure 3.5B) indicating HSF2's dependence on HSF1 for binding. We tested whether the reliance of HSF2 on HSF1 for DNA binding upon HS is a general feature or limited to *Hsps*. We took the union of all HSF2 peaks in WT cells and then quantified HSF2 ChIP-seq signal around those peaks in WT and *Hsf1*$^{-/-}$ cells. In WT cells, HSF2 binding gradually

Figure 3.5. HSF2 ChIP-seq in WT and *Hsf1*$^{-/-}$ MEFs
(A) Experimental set-up, ChIP-seq libraries in WT MEFs were made in duplicates with chromatin immunoprecipitated with non-specific IgG and a HSF2 specific antibody at NHS, 12'HS, and 60'HS. Additionally, ChIP-seq libraries were also made with the HF2 antibody in *Hsf1*$^{-/-}$ MEFs.
(B) Screenshot of *Hsph1* gene shows HSF2 ChIP-seq read density in WT and *Hsf1*$^{-/-}$ MEFs. The scale on the y-axis represents the relative number of ChIP-seq tags under the peaks.
(C) Screenshot of *Hsph1* gene shows HSF1 ChIP-seq (previously published) read density in WT and *Hsf1*$^{-/-}$ MEFs. The scale on the y-axis represents the relative number of ChIP-seq tags under the peaks.
(D) Composite profile of HSF2 ChIP-seq signal in WT MEFs (left panel) and *Hsf1*$^{-/-}$ MEFs (right panel) around all HSF2 peaks in WT MEFs.

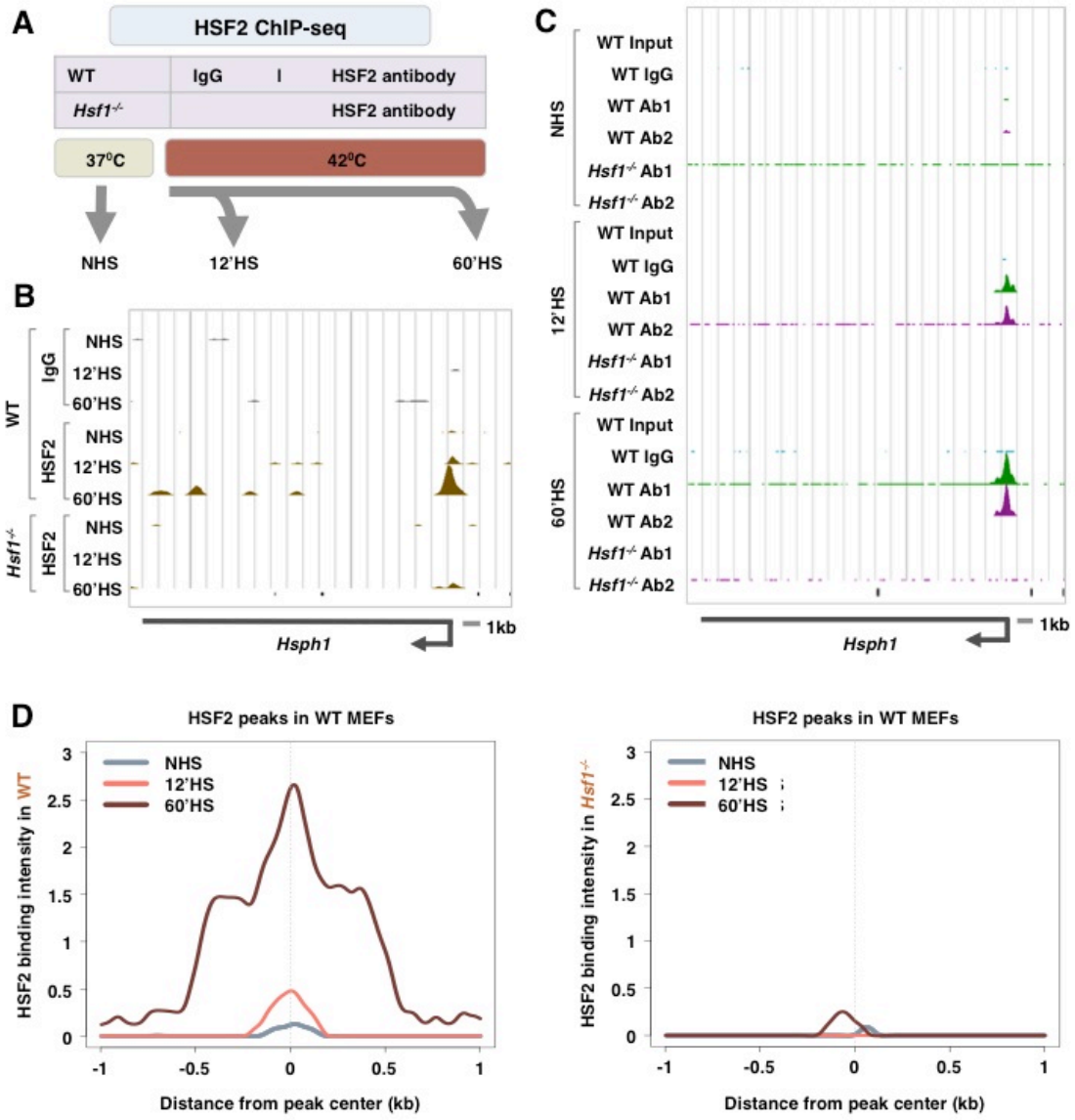

A
HSF2 ChIP-seq
WT
IgG
HSF2 antibody
Hsf1-/-
HSF2 antibody
37°C
42°C
NHS
12'HS
60'HS
B
WT
IgG
HSF2
Hsf1-/-
NHS
12'HS
60'HS
Hsph1
1kb
C
NHS
12'HS
60'HS
WT Input
WT IgG
WT Ab1
WT Ab2
Hsf1-/- Ab1
Hsf1-/- Ab2
Hsph1
1kb
D
HSF2 peaks in WT MEFs
HSF2 binding intensity in WT
NHS
12'HS
60'HS
Distance from peak center (kb)
HSF2 peaks in WT MEFs
HSF2 binding intensity in Hsf1-/-
NHS
12'HS
60'HS
Distance from peak center (kb)

Table 3.3. Sequencing depth & alignment statistics of ChIP-seq libraries made with HSF2 antibody or IgG.

Library	Sequenced reads	Adapter dimers	Unaligned	Uniquely aligned (mm10)
WT_NHS_IgG	31,175,847	3,395	611,608	30,560,844
WT_12'HS_IgG	20,594,110	4,019	445,030	20,145,061
WT_60'HS_IgG	42,710,902	2,851	1,164,766	41,543,285
WT_NHS_HSF2	45,864,673	3,588	888,834	44,972,251
WT_12'HS_HSF2	19,904,271	2,683	1,051,611	18,849,977
WT_60'HS_HSF2	16,100,424	2,975	1,081,322	15,016,127
Hsf1KO_NHS_IgG	13,825,759	4,717	1,154,664	12,666,378
Hsf1KO_NHS_HSF2	22,189,379	2,735	999,132	21,187,512
Hsf1KO_12'HS_HSF2	39,710,601	16,583	1,296,544	38,397,474
Hsf1KO_60'HS_HSF2	16,754,781	2,241	986,932	15,765,608

increases upon HS and exhibit prominent binding by 60 min of HS (Figure 3.5D, left panel). However, HSF2 binding is severely impaired in *Hsf1*$^{-/-}$ cells (Figure 3.5D, right panel). Together, these observations indicate that the dependence of HSF2 on HSF1 for DNA binding during HS is a general feature. At certain genes, HSF2 is known to heterotrimerize with HSF1 for binding (Sandqvist et al., 2009), and the requirement of HSF1 for HSF2 binding to the *Hsp70* promoter was previously reported (Ostling et al., 2007), however, HSF1's role in guiding HSF2 to the HS-specific genomic loci was not appreciated as a general rule.

Because HSF2 binding is virtually absent in *Hsf1*$^{-/-}$ cells, we focused HSF2 binding analysis on WT cells. We examined HSF2 binding on two class of genes: HSF2-dependent upregulated genes, an example of which is shown in Figure 3.6A, left panel, and HSF2-dependent downregulated genes, example shown in Figure 3.6A, right panel. HSF2 binding was absent from all HSF2-dependent upregulated genes and HSF2 binds to only 4% of the HSF2-dependent downregulated genes (Figure 3.6B). These results indicate that both HSF2-dependent transcription upregulation and downregulation during HS is a secondary effect of HSF2 likely mediated through other factors.

Transcription repression is mediated by inhibition of promoter-proximal pause release

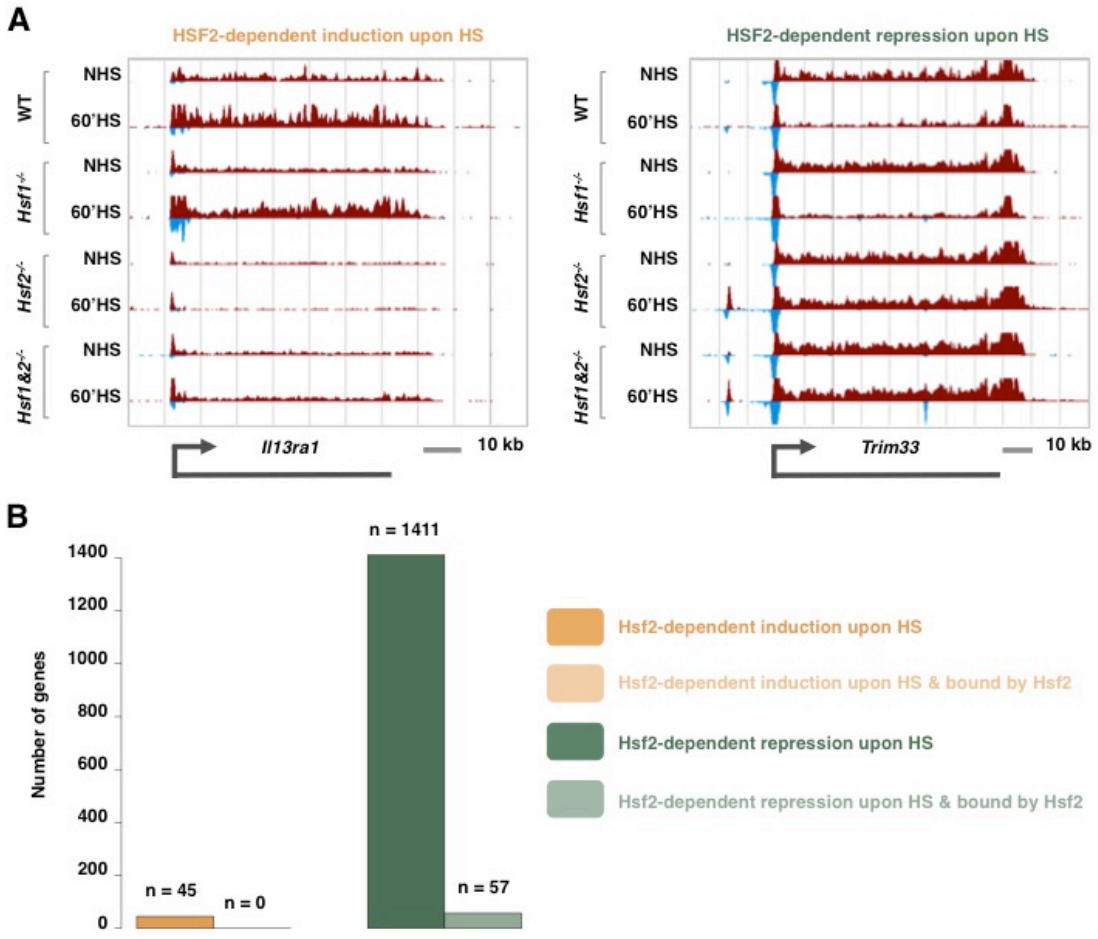


Figure 3.6. HSF2-dependent HS-regulated genes are not bound by HSF2
(A) Screenshots of *Il13ra1* gene representing HSF2-dependent induction during HS (left panel) and *Trim33* gene representing HSF2-dependent repression (right panel). PRO-seq density in sense and antisense direction is shown in red and blue respectively. Small vertical black bars at the bottom of PRO-seq tracks represent the genomic regions that do not map uniquely at 36bp resolution.
(B) Total numbers of genes regulated in HSF2-dependent manner are shown in darker shades (induced in orange and repressed in green) and the numbers of genes bound by HSF2 within those catagories are shown in lighter shades.

The lack of HSF2 severely reduces the pool of downregulated genes during HS, from ~8000 in WT and *Hsf1*$^{-/-}$ to ~5000 in *Hsf2*$^{-/-}$ and *Hsf1&2*$^{-/-}$ cells. Nevertheless, 5000 genes still account for a third of actively transcribing genes. We previously showed that the global downregulation of transcription in WT and *Hsf1*$^{-/-}$ cells is caused by inhibition of Pol II release from promoter-proximal pause (Mahat et al., 2016). We found the same regulatory mechanism being responsible for global downregulation of genes in *Hsf2*$^{-/-}$ and *Hsf1&2*$^{-/-}$ (Figure 3.7A). The increase in paused Pol II is confined to the first few hundred bps from the TSS (Figure 3.7B). This finding further establishes the promoter-proximal pause release as the key regulatory step in transcription regulation during HS.

HSF2 does not compensate for the loss of HSF1 during HSR

Our recent work established the dependence of transcription induction of heat-inducible *Hsps* on HSF1. However, the interplay of HSF1 and HSF2 in induction of *Hsps* is unclear. HSF1 and HSF2 can form heterotrimer and we showed earlier that HSF2 binds to the promoter of heat-inducible *Hsps* in WT MEFs (Figure 3.5B). To dissect the contribution of HSF2 in induction of *Hsps*, we compared the transcriptional change in major *Hsps* in *Hsf2*$^{-/-}$ and *Hsf1&2*$^{-/-}$ with WT and *Hsf1*$^{-/-}$ cells. The profile of transcriptional change in *Hsf2*$^{-/-}$ looks very similar to WT – heat-inducible *Hsps* are robustly induced in both cell types (Figure 3.8). Similarly, the heat-inducible *Hsps* are not induced in *Hsf1*$^{-/-}$

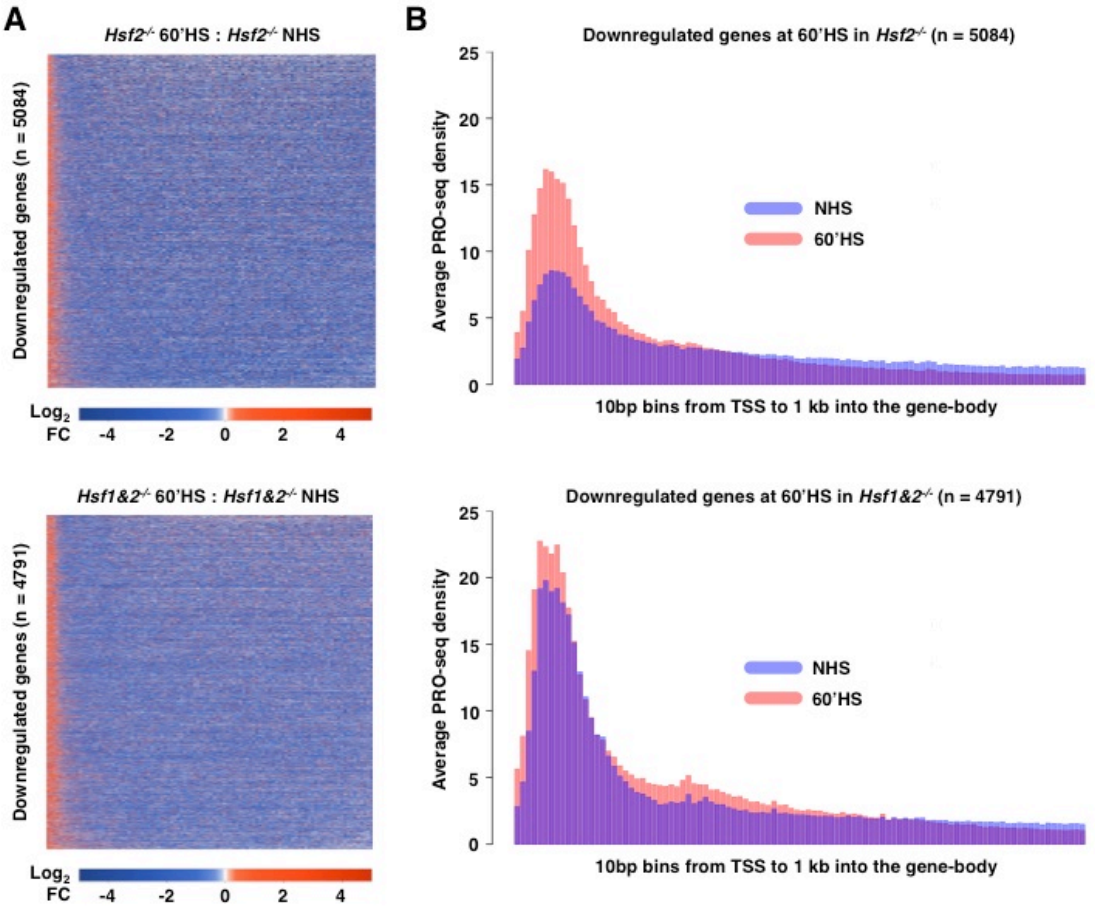


Figure 3.7. Transcription repression is mediated by inhibition of promoter-proximal pause release

(A) Heatmap of PRO-seq density fold change from 60'HS to NHS for significantly downregulated genes at 60'HS in *Hsf2*$^{-/-}$ (top) and *Hsf1&2*$^{-/-}$ MEFs (bottom). Each row represents a gene, scaled to same length and divided into 100 bins, from TSS up to polyA site for genes shorter than 24 kb and up to 24 kb for genes longer than 24kb.

(B) Histogram of average PRO-seq density in 10 bp bins from TSS to 1 kb into the gene-body for significantly downregulated genes at 60'HS in *Hsf2*$^{-/-}$ (top) and *Hsf1&2*$^{-/-}$ MEFs (bottom).

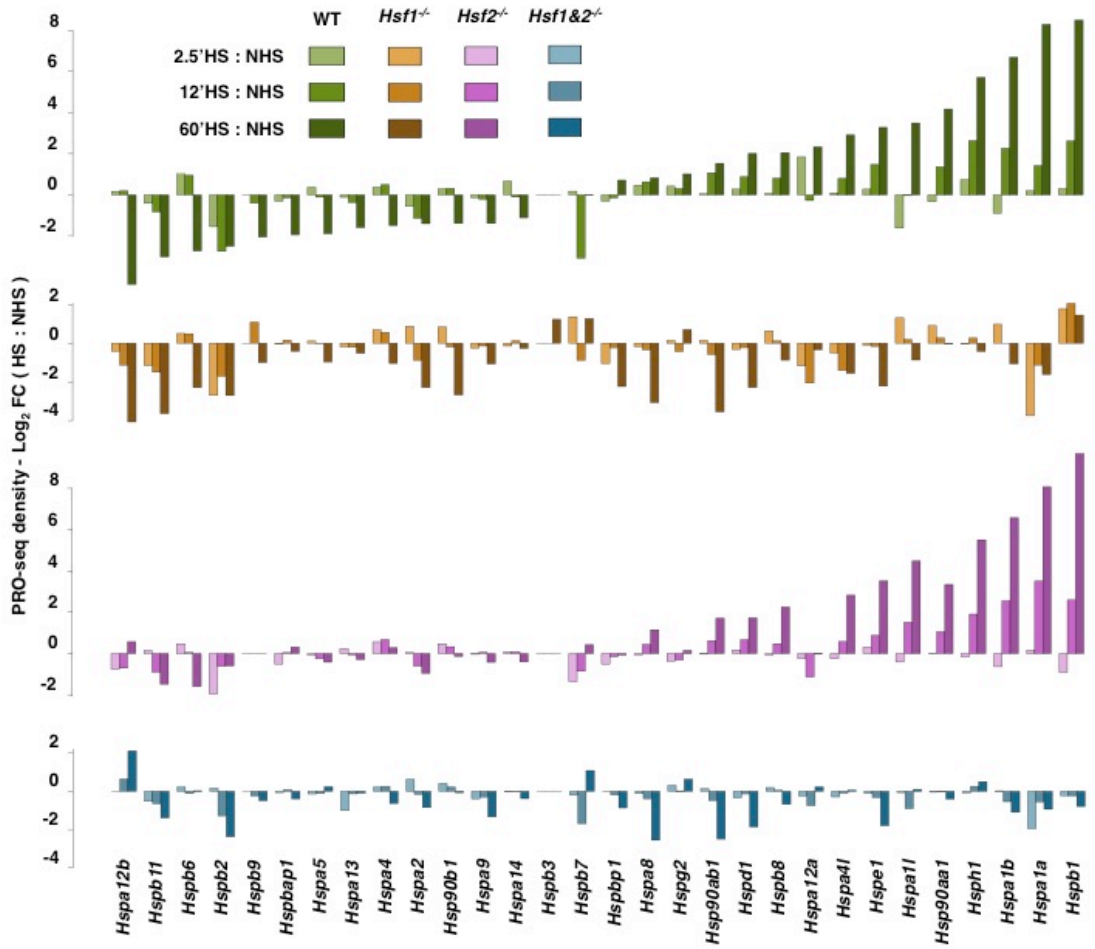


Figure 3.8. HSF2 does not compensate for the loss of HSF1 during HSR
PRO-seq density change upon HS in 30 *Hsps* in WT, *Hsf1*$^{-/-}$, *Hsf2*$^{-/-}$, and *Hsf1&2*$^{-/-}$ MEFs (top to bottom) at 2.5, 12, and 60 min of HS. The genes are ordered from left to right by increasing fold change at WT 60'HS.

and *Hsf1&2*$^{-/-}$, emphasizing their dependence on HSF1 (Figure 3.8). The similarity in the regulation of *Hsps* in *Hsf1*$^{-/-}$ and *Hsf1&2*$^{-/-}$ further indicates that HSF2 does not compensate for the lack of HSF1.

Although the regulation of *Hsps* look very similar between WT and *Hsf2*$^{-/-}$ cells, the extent of upregulation is higher and downregulation is lower in *Hsf2*$^{-/-}$ compared to WT cells. This suggests that the binding of HSF2 to the *Hsps* promoters, possibly as a heterotrimer with HSF1 in HSF1-dependent manner, likely compromises the induction potential of HSF1. Again, this role of HSF2 in attenuating the upregulation of *Hsps* further bolsters our hypothesis of HSF2 being more of a repressor than inducer of transcription during HS.

Discussion

The discovery of HSF2 (Sarge et al., 1991; Schuetz et al., 1991) – a paralog of HSF1 that shares similarity to HSF1 in sequence composition, protein structure, and DNA substrate for binding – has incited long-standing questions on its functional relevance and evolutionary discretion for its selection. It is exclusively present in vertebrates but whether it complements HSF1 or compromises it remained poorly understood. Past studies on a few *Hsp* genes using biochemical, optical, and genetic tools have provided valuable information on HSF2's role. However, the generality of such anecdotal evidence required comprehensive studies. In this study, we used genome-wide approaches to explore some unresolved aspects of HSF2

biology. We quantified global nascent transcription at multiple time-points during HSR in *Hsf2*$^{-/-}$ and *Hsf1&2*$^{-/-}$ cells using PRO-seq – a genomic assay with high sensitivity and base pair resolution. Comparing the transcriptional profile change in these two cell types and with our previously reported study on WT and *Hsf1*$^{-/-}$ cells, we found 45 genes induced in HSF2-dependent manner during HSR, but no *Hsps*. On the contrary, HSF2 was responsible for more than a third of repressed genes in WT and *Hsf1*$^{-/-}$ cells. To distinguish the primary from secondary role of HSF2 on transcription regulation, we generated HSF2 ChIP-seq data. Binding landscape of HSF2 revealed that the HSF2-dependent repressed genes were likely mediated through a secondary factor, as they lacked direct HSF2 binding. Furthermore, we showed that the inducible binding of HSF2 at some *Hsp* promoters during HS is dependent on HSF1, likely as a heterotrimer, and the hitchhiked HSF2 marginally attenuates the activating potential of HSF1.

Our previous study of HSR in WT and *Hsf1*$^{-/-}$ MEFs identified many genes that were similarly induced in these cell types. HSF2 is a conceivable candidate that could compensate for the lack of HSF1 in controlling the commonly regulated genes in WT and *Hsf1*$^{-/-}$. Despite the identical DNA binding sites and similar DNA binding domain, we learned from this study that HSF2 does not compensate for the absence of HSF1 for induction of commonly regulated genes in WT and *Hsf1*$^{-/-}$ MEFs, nor does it for *Hsps*, indicating non-redundant roles of this ubiquitously-expressed HSF1 paralog.

This observation warrants an explanation for evolutionary favorability of multiple HSFs in vertebrates. One likely explanation may reside beyond the sphere of HSR. Perhaps the HSF2 that resulted from duplication of HSF1 prior to the vertebrate radiation structurally evolved and specialized on non-critical constitutive cellular functions. DNA-proximal domains of HSF1 and HSF2 share similar features, however, DNA-distal parts such as trimerization and activation domains exhibit different characteristics thereby driving diverse post-translational modifications and interacting partners (Jaeger et al., 2016). Constitutive HSF2 function is implicated in brain development (Kallio et al., 2002) and fetal alcohol syndrome (Fatimy et al., 2014). HSF2 is known to suppress prostate cancer invasion (Björk et al., 2016) and mutations in HF2 is associated with idiopathic azoospermia (Mou et al., 2013; Wilkerson et al., 2008). Therefore, the limited role of HSF2 in transcription induction during HSR should not be confused for its lack of significance as it may have specialized in non-stress-related cellular functions, many of which remain to be shown.

HSF2 clearly played a major role in transcription repression during HSR. Anecdotal evidence of HSF2's role in repression has been previously documented – it is known to repress HSR in mitotic cells, (Elsing et al., 2014), but the dominant repression of 1411 genes during HS has not been reported before in interphase cells. Lack of HSF2 binding at these promoters supports a model where either inducer of these genes during HS is constitutively

repressed by HSF2 or their repressor is induced by HSF2 upon stress. Alternatively, HSF2 could bind to distal regulatory elements and control transcription through long-range mediated interaction, although this scenario is rather unlikely due to lack of HSF2-bound intergenic and intragenic genomic sites (data not shown). The use of knock-out cells in this study, where the cells have likely compensated for the constitutive lack of HSF1 or HSF2, complicates our attempt to identify underlying mechanism of their function. In future, novel tools and approaches that can rapidly and robustly impair the function of specific domains or the entire protein can be used to directly probe the primary role of HSF2 during HSR.

REFERENCES

Ahn, S.G., Liu, P.C., Klyachko, K., Morimoto, R.I., and Thiele, D.J. (2001). The loop domain of heat shock transcription factor 1 dictates DNA-binding specificity and responses to heat stress. Genes & Development *15*, 2134–2145.

Akerfelt, M., Henriksson, E., Laiho, A., Vihervaara, A., Rautoma, K., Kotaja, N., and Sistonen, L. (2008). Promoter ChIP-chip analysis in mouse testis reveals Y chromosome occupancy by HSF2. Proceedings of the National Academy of Sciences *105*, 11224–11229.

Akerfelt, M., Morimoto, R.I., and Sistonen, L. (2010). Heat shock factors: integrators of cell stress, development and lifespan. Nat Rev Mol Cell Biol *11*, 545–555.

Alastalo, T.-P., Hellesuo, M., Sandqvist, A., Hietakangas, V., Kallio, M., and Sistonen, L. (2003). Formation of nuclear stress granules involves HSF2 and coincides with the nucleolar localization of Hsp70. J. Cell. Sci. *116*, 3557–3570.

Björk, J.K., Åkerfelt, M., Joutsen, J., Puustinen, M.C., Cheng, F., Sistonen, L., and Nees, M. (2016). Heat-shock factor 2 is a suppressor of prostate cancer invasion. Oncogene *35*, 1770–1784.

Elsing, A.N., Aspelin, C., Björk, J.K., Bergman, H.A., Himanen, S.V., Kallio, M.J., Roos-Mattjus, P., and Sistonen, L. (2014). Expression of HSF2 decreases in mitosis to enable stress-inducible transcription and cell survival. The Journal of Cell Biology *206*, 735–749.

Fatimy, El, R., Miozzo, F., Le Mouël, A., Abane, R., Schwendimann, L., Sabéran-Djoneidi, D., de Thonel, A., Massaoudi, I., Paslaru, L., Hashimoto-Torii, K., et al. (2014). Heat shock factor 2 is a stress-responsive mediator of neuronal migration defects in models of fetal alcohol syndrome. EMBO Mol Med *6*, 1043–1061.

Jaeger, A.M., Pemble, C.W., Sistonen, L., and Thiele, D.J. (2016). Structures

of HSF2 reveal mechanisms for differential regulation of human heat-shock factors. Nat. Struct. Mol. Biol. *23*, 147–154.

Kallio, M., Chang, Y., Manuel, M., Alastalo, T.-P., Rallu, M., Gitton, Y., Pirkkala, L., Loones, M.-T., Paslaru, L., Larney, S., et al. (2002). Brain abnormalities, defective meiotic chromosome synapsis and female subfertility in HSF2 null mice. Embo J. *21*, 2591–2601.

Kwak, H., Fuda, N.J., Core, L.J., and Lis, J.T. (2013). Precise maps of RNA polymerase reveal how promoters direct initiation and pausing. Science *339*, 950–953.

Lecomte, S., Desmots, F., Le Masson, F., Le Goff, P., Michel, D., Christians, E.S., Le Dréan, Y., and Le Dréan, Y. (2010). Roles of heat shock factor 1 and 2 in response to proteasome inhibition: consequence on p53 stability. Oncogene *29*, 4216–4224.

Lecomte, S., Reverdy, L., Le Quément, C., Le Masson, F., Amon, A., le Goff, P., Michel, D., Christians, E., and le Dréan, Y. (2013). Unraveling complex interplay between heat shock factor 1 and 2 splicing isoforms. PLoS ONE *8*, e56085.

Mahat, D.B., Salamanca, H.H., Duarte, F.M., Danko, C.G., and Lis, J.T. (2016). Mammalian Heat Shock Response and Mechanisms Underlying Its Genome-wide Transcriptional Regulation. Molecular Cell *62*, 63–78.

Manuel, M., Rallu, M., Loones, M.-T., Zimarino, V., Mezger, V., and Morange, M. (2002). Determination of the consensus binding sequence for the purified embryonic heat shock factor 2. Eur. J. Biochem. *269*, 2527–2537.

McMillan, D.R., Christians, E., Forster, M., Xiao, X., Connell, P., Plumier, J.C., Zuo, X., Richardson, J., Morgan, S., and Benjamin, I.J. (2002). Heat Shock Transcription Factor 2 Is Not Essential for Embryonic Development, Fertility, or Adult Cognitive and Psychomotor Function in Mice. Mol. Cell. Biol. *22*, 8005–8014.

Mou, L., Wang, Y., Li, H., Huang, Y., Jiang, T., Huang, W., Li, Z., Chen, J.,

Xie, J., Liu, Y., et al. (2013). A dominant-negative mutation of HSF2 associated with idiopathic azoospermia. Hum. Genet. *132*, 159–165.

Neef, D.W., Jaeger, A.M., and Thiele, D.J. (2011). Heat shock transcription factor 1 as a therapeutic target in neurodegenerative diseases. Nat Rev Drug Discov *10*, 930–944.

Neudegger, T., Verghese, J., Hayer-Hartl, M., Hartl, F.U., and Bracher, A. (2016). Structure of human heat-shock transcription factor 1 in complex with DNA. Nat. Struct. Mol. Biol. *23*, 140–146.

Ostling, P., Björk, J.K., Roos-Mattjus, P., Mezger, V., and Sistonen, L. (2007). Heat shock factor 2 (HSF2) contributes to inducible expression of hsp genes through interplay with HSF1. J. Biol. Chem. *282*, 7077–7086.

Sandqvist, A., Björk, J.K., Akerfelt, M., Chitikova, Z., Grichine, A., Vourc'h, C., Jolly, C., Salminen, T.A., Nymalm, Y., and Sistonen, L. (2009). Heterotrimerization of heat-shock factors 1 and 2 provides a transcriptional switch in response to distinct stimuli. Mol. Biol. Cell *20*, 1340–1347.

Sarge, K.D., Zimarino, V., Holm, K., Wu, C., and Morimoto, R.I. (1991). Cloning and characterization of two mouse heat shock factors with distinct inducible and constitutive DNA-binding ability. Genes & Development *5*, 1902–1911.

Schuetz, T.J., Gallo, G.J., Sheldon, L., Tempst, P., and Kingston, R.E. (1991). Isolation of a cDNA for HSF2: evidence for two heat shock factor genes in humans. Proc. Natl. Acad. Sci. U.S.a. *88*, 6911–6915.

Vihervaara, A., Sergelius, C., Vasara, J., Blom, M.A.H., Elsing, A.N., Roos-Mattjus, P., and Sistonen, L. (2013). Transcriptional response to stress in the dynamic chromatin environment of cycling and mitotic cells. Proceedings of the National Academy of Sciences *110*, E3388–E3397.

Whitesell, L., and Lindquist, S. (2009). Inhibiting the transcription factor HSF1 as an anticancer strategy. Expert Opin. Ther. Targets *13*, 469–478.

Wilkerson, D.C., Murphy, L.A., and Sarge, K.D. (2008). Interaction of HSF1 and HSF2 with the Hspa1b promoter in mouse epididymal spermatozoa. Biol. Reprod. *79*, 283–288.

Wilkerson, D.C., Skaggs, H.S., and Sarge, K.D. (2007). HSF2 binds to the Hsp90, Hsp27, and c-Fos promoters constitutively and modulates their expression. Cell Stress Chaperones *12*, 283–290.

Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W., et al. (2008). Model-based analysis of ChIP-Seq (MACS). Genome Biol. *9*, R137.

CHAPTER 4

TRANSCRIPTIONAL RECOVERY AFTER HEAT SHOCK RESPONSE

Summary

Heat shock response is a cellular defense mechanism against heat stress and its activation can protect cells from stress-induced apoptosis. Cells that have recovered from heat stress perform better against subsequent exposure to such stress. This acquired thermotolerance is prevalent across wide spectrum of life. However, the molecular mechanism for this evolutionary beneficial response is poorly understood. Heat shock proteins (HSPs), which are molecular chaperones, are known to be critical for thermotolerance, however, the significance of heat shock transcription factor that drive the expression of HSPs remains disputed. Here we probed the transcriptional control of thermotolerance at genome-wide scale. We found a surprising lack of transcriptional changes at genes and enhancers after recovery from HS. Nevertheless, a subset of anti-apoptotic genes that respond slowly to heat stress shows an increase in RNA polymerase II level at promoter-proximal sites upon recovery. This increase pausing may serve to prime the rapid induction of anti-apoptotic genes during re-exposure to heat stress and could explain the survival benefits of acquired thermotolerance by providing time-advantage for rapid deployment of anti-apoptotic factors.

Introduction

Heat shock response (HSR) is widely studied for its role in maintaining protein homeostasis under proteotoxic conditions. Although some aspects of this protective mechanism remain to be understood, the field of stress biology has made significant progress in elucidating the stress-responsive genes, mechanisms involved in their regulation, and the regulatory factors driving the response. The primary hallmark of HSR is the fast and furious induction of heat shock protein genes (*Hsps*) by the heat shock transcription factor-1 (HSF1). While severe or sustained subjection to elevated temperature is detrimental to cells, successful HSR restores proteostasis and prevents cells from stress-induced apoptosis. Once cells overcome the first exposure to heat stress, they tend to perform significantly better when they re-encounter exposure to heat stress again. This acquired ability to withstand heat stress due to mild conditioning pretreatments is known as thermotolerance.

Although thermotolerance is a widely-shared acquired phenotype from archaea to fungi to plants to animals (Kregel, 2002; Queitsch et al., 2000; Singer and Lindquist, 1998; Trent, 1996), the precise mechanisms of its establishment and employment are not fully understood. Previous works on thermotolerance in various species also have provided support for contradictory mechanisms. HSP101 is critical for theromotolerance in Arabidopsis, while HSP104 is critical in yeast (Lindquist and Kim, 1996; Queitsch et al., 2000). In yeast, neither expression of HSP104 nor

thermotolerance is dependent on HSF (Lindquist and Kim, 1996), while in mammals, thermotolerance is dependent on HSF1 (Baird et al., 2014; McMillan et al., 1998; Zhang et al., 2002). These observations collectively suggest a likely role of HSPs in a cell's ability to tolerate heat stress despite HSF-independent and HSF1-dependent induction of chaperone proteins in yeast and mammals respectively. Most of the past studies on thermotolerance were done shortly after the conditioning treatment, which raises the possibility that their ability to tolerate heat stess is due to higher levels of HSPs generated during the conditioning treatment and thus conferring thermotolerance. However, proteomic analysis of yeasts under prolonged thermal stress, presumably thermotolerant, found significant difference in protein composition from that of yeast undergoing heat shock response (HSR) (Shui et al., 2015) suggesting a major departure from HS in regulatory programs during thermotolerance. If altered expression of certain genes is the source of thermotolerance, whether it is higher expression of HSPs or change in expression of a different set of genes, we would expect either transcription, steady-state mRNA level, translation, or a combination of these should deviate from the basal level.

In this study, we tested whether earlier exposure to heat stress modifies transcription regulatory switches and if those modifications contribute to thermotolerance. We used a highly sensitive genome-wide PRO-seq assay to measure transcriptionally-engaged RNA Polymerase II (Pol II) in mouse

embryonic fibroblasts before and after recovery from HS. We found a complete recovery of transcriptional changes at HS-regulated genes and enhancers. However, we observed a higher level of promoter-proximal paused Pol II at many genes upon recovery form HS. When we probed genes with increased pausing after recovery form HS that were also induced during HS, we found that those genes encode anti-apoptotic factors and were late responders during the first HS. It is possibility that the higher pause in late-responding anti-apoptotic genes can shorten the non-responsive period by accelerating the induction kinetics of these genes when they re-encounter HS again. Perhaps this enhanced responsiveness of anti-apoptotic genes is one of many paths to thermotolerance.

Materials and methods:

Cell culture

Mouse embryonic fibroblasts were grown in 150mm TC-treated and gamma irradiated cell culture dish in Dulbecco's Modified Eagle Medium supplemented with 10% heat inactivated fetal bovine serum (v/v) and 1% Penicillin Streptomycin (v/v) at 37^0C with 5% CO_2 and 90% humidity.

HS and recovery from HS

HS was performed at 42^0C incubator with 5% CO_2 and 95% humidity. Tissue culture plates at ~80% confluency were moved from 37^0C incubator

directly to 42^0C incubator. Unlike our previous work, these cells were not instantaneously heat shocked. We allowed 15 minutes for temperature in tissue culture plates to equilibrate to 42^0C and then continued for full 60 minutes of HS after equilibration.

For short recovery from HS, cells were placed back into 37^0C for 4 hrs. For long recovery from HS, cells were place back into 37^0C for 4 hours and were split 1:8 to let it recover for another 44 hours (total 48 hours). Based on the doubling time, we estimated the long recovery from HS allowed ~2 cell divisions.

PRO-seq library preparation and data analysis

PRO-seq library preparation and data analysis were performed as described previously (Mahat et al., 2016).

Results

HS-induced transcriptional changes in genes recover completely upon reverting to normal temperature

To test the transcriptional status of HS-regulated genes after recovery from HS, we used PRO-seq to measure genome-wide change in transcription. The information on transcriptionally engaged Pol II density obtained from PRO-seq provides direct and quantitative measure of transcription level. We made PRO-seq libraries in mouse embryonic fibroblasts at four different

Figure 4.1. Transcriptional status of HS-regulated genes after recovery from HS

(A) Experimental set-up, PRO-seq assay was performed in nuclei isolated from WT MEFs at prior to HS (37^0C - NHS), during HS (1 hour at 42^0C – 1hr-HS), short recovery after HS (1 hr at 42^0C followed by 4 hrs at 37^0C - 1hr-HS_4hr-Recovery), and long recovery after HS (1 hr at 42^0C followed by 48 hrs at 37^0C - 1hr-HS_48hr-Recovery).

(B) Screenshots of upregulated genes during HS (left panels) and downregulated genes during HS (right panels) showing PRO-seq density before, during, and after recovery from HS. PRO-seq density in sense and antisense direction is shown in orange and green respectively. Small vertical black bars at the bottom of PRO-seq tracks represent the genomic regions that do not map uniquely at 36bp resolution.

(C) 'Minus-average' (MA) plots show change in PRO-seq density in all genes between NHS and 1hr-HS (left panel), 1hr-HS_4hr-Recovery (mid panel), and 1hr-HS_48hr-Recovery (right panel) in WT MEFs. Significantly upregulated genes are shown in green, significantly downregulated ones are shown in orange, and the rest are shown in grey.

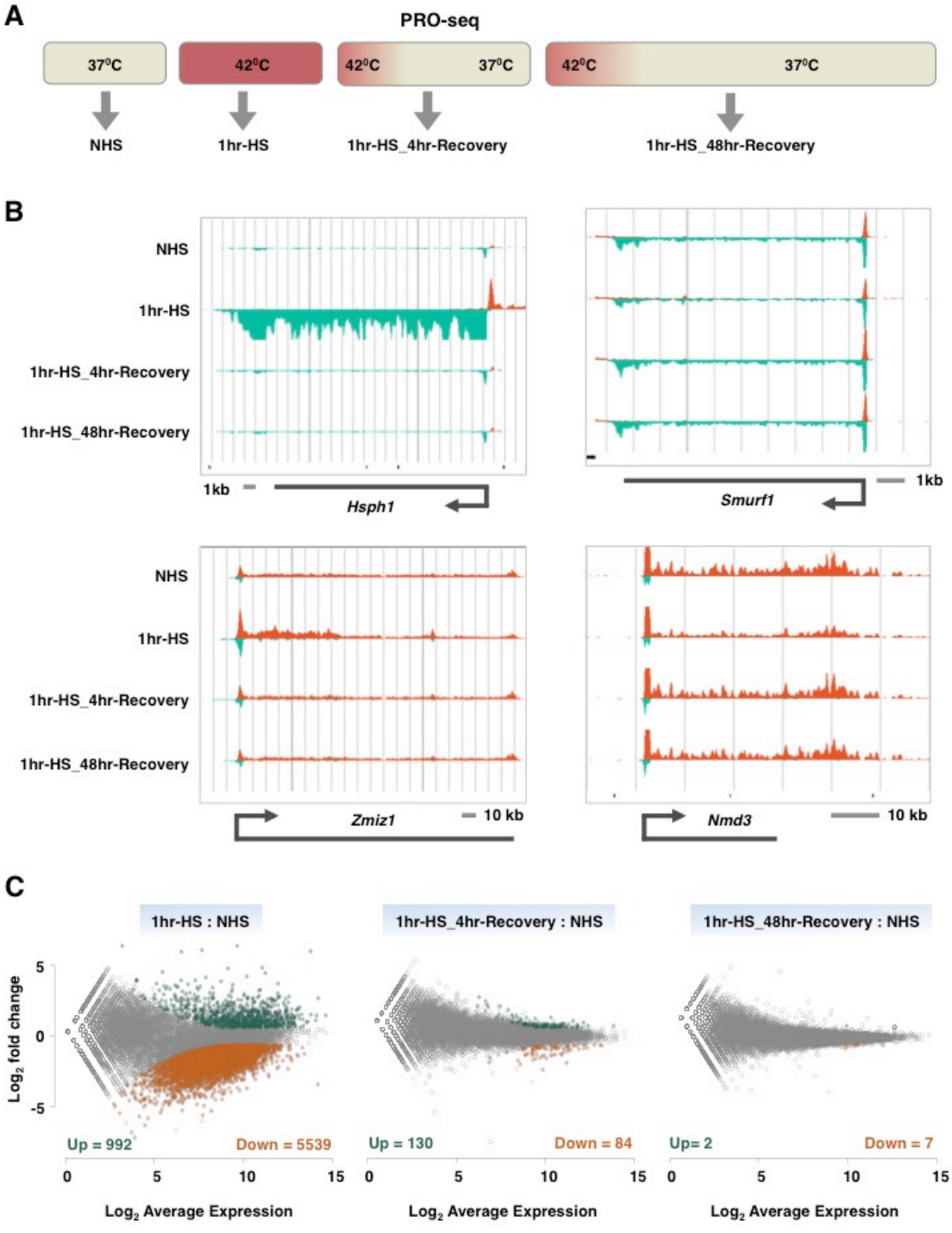

A
PRO-seq
37°C
42°C
42°C
37°C
42°C
37°C
NHS
1hr-HS
1hr-HS_4hr-Recovery
1hr-HS_48hr-Recovery
B
NHS
1hr-HS
1hr-HS_4hr-Recovery
1hr-HS_48hr-Recovery
1kb
Hsph1
Smurf1
1kb
Zmiz1
10 kb
Nmd3
10 kb
C
1hr-HS : NHS
1hr-HS_4hr-Recovery : NHS
1hr-HS_48hr-Recovery : NHS
Log2 fold change
5
0
-5
Up = 992
Down = 5539
Up = 130
Down = 84
Up= 2
Down = 7
0
5
10
15
Log2 Average Expression

Table 4.1. Sequencing depth and alignment statistics of PRO-seq libraries before, during, and after recovery from HS.

Library	Sequenced	Un-aligned	Multi-aligned	Uniquely aligned (mm10)
NHS_BR1	99,557,605	46.57%	35.67%	14,665,417
NHS_BR2	51,324,882	40.13%	34.88%	10,019,742
1hr-HS _BR1	39,204,942	26.09%	38.94%	12,351,604
1hr-HS _BR2	37,285,781	39.12%	31.68%	8,045,648
1hr-HS_4hr-Recovery _BR1	42,695,172	37.46%	39.24%	8,880,102
1hr-HS_4hr-Recovery _BR2	57,075,503	52.53%	30.13%	5,870,107
1hr-HS_48hr-Recovery _BR1	39,152,023	43.65%	37.22%	6,521,228
1hr-HS_48hr-Recovery _BR2	27,108,022	46.64%	32.52%	4,064,154

conditions: prior to HS (37^{0}C - NHS), during HS (60 min at 42^{0}C – 1hr-HS), short recovery after HS (1 hr at 42^{0}C followed by 4 hrs at 37^{0}C - 1hr-HS_4hr-Recovery), and long recovery after HS (1 hr at 42^{0}C followed by 48 hrs at 37^{0}C - 1hr-HS_48hr-Recovery) (Figure 4.1A and Table 4.1). The choice of 4 and 48 hours of recovery after HS allows observation of transcription recovery kinetics at short- and long-term respectively. Comparison of PRO-seq profiles show that the genes induced by HS, both HSPs and non-HSPs, transcriptionally recover to basal level by 4 hours at normal temperature (Figure 4.1B, left panels). Similarly, transcription of genes repressed by HS also recovers to basal level within 4 hours in recovery temperature (Figure 4.1B, right panels). At genome-wide level, transcription of most genes restores by 4 hours in recovery temperature, and by 48 hours, virtually all genes recover transcriptionally (Figure 4.1C). Genes that are significantly upregulated or downregulated even after 4 hours of recovery are the genes that were most robustly upregulated and downregualted respectively at 60 min of HS. These and other genes that are still called significantly changed after 48 hours of recovery lie at the border of unchanged and changed and are likely result of higher statistical significance due to high read density. Overall, our analyses of PRO-seq profiles indicate that transcriptional changes in genes during HS completely restore to basal level upon recovery from HS.

Transcriptional changes at enhancers also recover from HS

Transcription is not limited to the stable-RNA-encoding genes. It is well established that many regions of the genome are transcribed (Djebali et al., 2012), including enhancers. Enhancers can act from a distance to regulate transcription of genes (Plank and Dean, 2014). Our previous study identified many genes that are regulated in HSF1-dependent manner but without detectable HSF1 binding at their promoters (Mahat et al., 2016) suggesting a possibility of distant-acting HSF1-bound enhancers regulating those genes. Thus, we hypothesized that the thermotolerance could partially be mediated by HS-induced enhancers that remain activate after HS recovery. The heightened activity of these enhancers prior to HS in recovered cells could provide time-advantage for rapid and robust activation of target genes when exposed to HS again. To test this hypothesis, we examined PRO-seq densities around the known HS-activated enhancers (Mahat et al., 2016). Active enhancers are characterized by divergent transcription that produce short unstable RNAs (Core et al., 2014; 2008; Seila et al., 2008), represented by divergent peaks in PRO-seq metagene profile. We looked at intergenic and intragenic HS-activated enhancers and found a complete recovery of transcription at both sets of enhancers (Figure 4.2). This observation disproves our hypothesis of constitutive activation of HS-activated enhancers upon recovery from stress and suggests a lack of enhancer role in thermotolerance of stress-recovered cells.

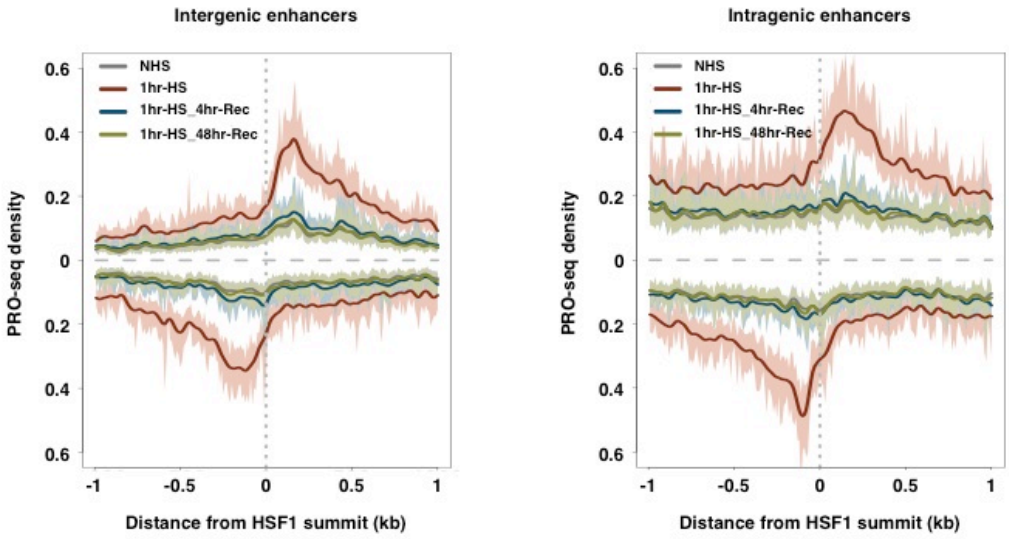


Figure 4.2. Transcriptional status at HS-activated enhancers after recovery from HS

PRO-seq density before, during, and at time points after recovery from HS in WT MEFs around the center of HS-activated enhancers at intergenic sites (left) and intragenic sites (right).

Increased pausing after recovery in late-induced genes during HS

The lack of transcriptional change at annotated genes and HS-activated enhancers after recovery from HS was at the very least surprising. Our imagination of transcriptional preparedness bestowing thermotolerance in stress-recovered cells, analogous to a city's preparedness in fire protection that was once engulfed in flames, lacked evidence. After ruling out two major candidates in our quest for identifying the genomic features responsible for thermotolerance, we focused on other areas of transcription that has potential to increase the reaction time of transcription induction. One genomic feature with such potential is promoter-proximal paused Pol II (Adelman and Lis, 2012). Segmentation genes and HS-inducible genes in *Drosophila* have high pausing (Duarte et al., 2016; Levine, 2011; Wang et al., 2007), and one major difference between early induced and late induced genes during mammalian HSR is also the level of Pol II pausing (Mahat et al., 2016). Therefore, we hypothesized that some genes may have increased Pol II pausing after recovery from HS priming those genes for faster than intrinsic induction kinetics during re-exposure to HS. To test this hypothesis, we measured the genome-wide Pol II pausing and found an increase in global pausing after HS recovery even though the gene-body transcription returns to basal level (Figure 4.3A). As promising as this observation was, there is a possibility that it could be a result of slow recovery of transcription at pause regions of HS-activated genes. To rule out this suspicion, we juxtaposed PRO-seq density

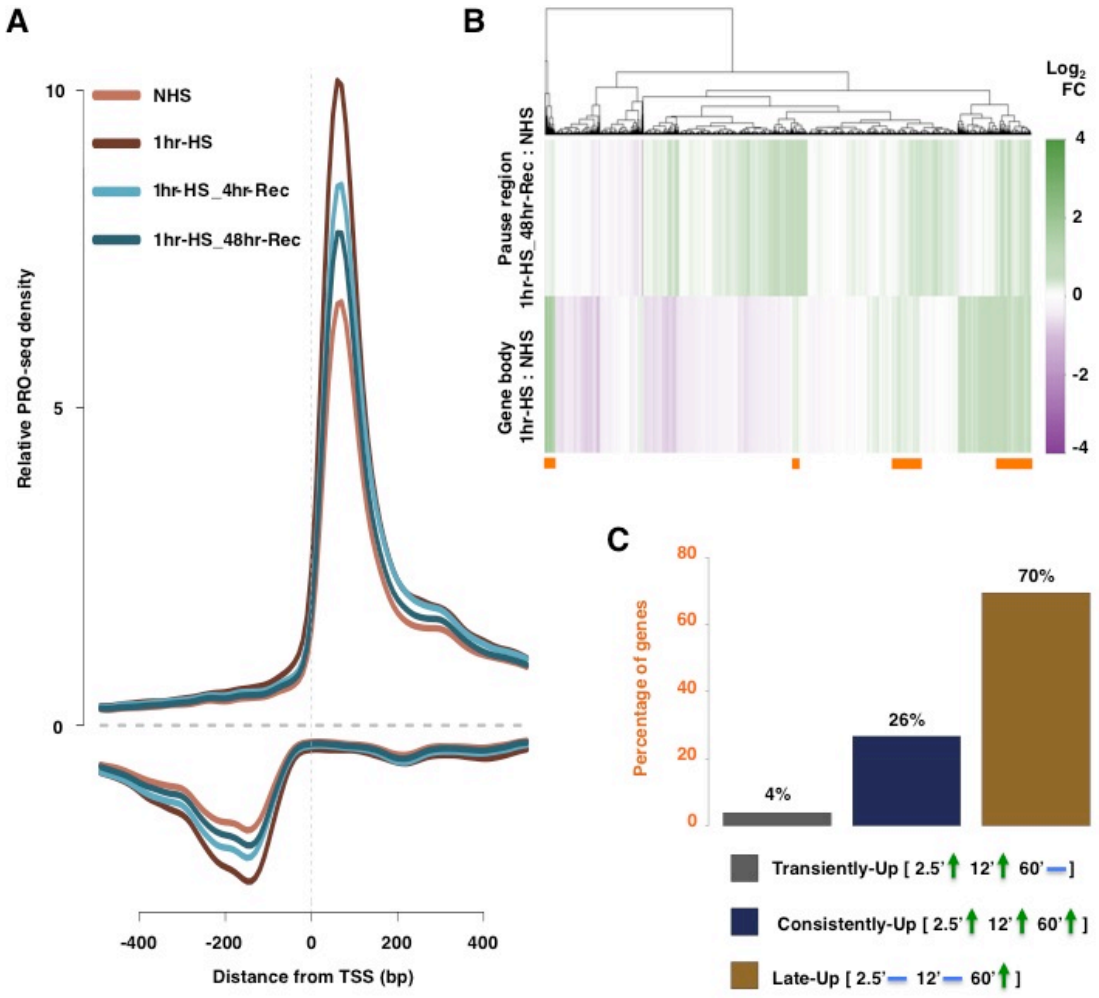


Figure 4.3. Increased promoter-proximal Pol II pausing after recovery from HS in a selected class of genes

(A) A meta-gene profile showing PRO-seq density around the TSS of all genes before, during, and at time points after recovery from HS. Positive value on y-axis represents transcription in sense strand and negative value represents transcription in anti-sense strand.

(B) Heatmap of PRO-seq density fold change during HS in the gene body region and after complete recovery from HS in pause region shown side-by-side. The orange bars at the bottom of heatmap indicate a special class of genes that have increased transcription during HS and also have higher pausing after recovery from HS.

(C) Barplot showing the percentage of genes that have increased transcription during HS and also have higher pausing after recovery from HS (shown as orange bars in Figure 4.3B) in three different classes based on kinetics of induction (2.5 min of HS, 12 min of HS, and 60 min of HS from our previously published work) during HS. The green arrows indicate increase in transcription at that time point and the blue bars represent no change in transcription.

fold-change in gene body at 60 min of HS with change in pause region after complete (48 hours) recovery from HS (Figure 4.3B). Interestingly, genes with an increase in reads in the pause region after HS recovery did not overlap well with genes with an increase in gene body reads during HS (Figure 4.3C); many genes with increased pausing after recovery were downregulated during HS. We speculated the genes that are upregulated during HS (and thus have pro-survival functions) and have higher pausing after HS recovery might have a role in thermotolerance (orange bars in Figure 4.3B). We examined the induction kinetics of these genes using data from our previous study (Mahat et al., 2016) and found most of these genes are induced late during HS (Figure 4.3C). These genes are enriched for anti-apoptotic functions according to gene ontology analysis (Mahat et al., 2016). Therefore, it is tempting to assume that the higher pausing in anti-apoptotic genes allows faster-than-intrinsic transcription induction of these genes when cells re-encounter HS. Such 'readiness' can enable early deployment of anti-apoptotic factors and likely boost cell's survival chances.

Discussion

Our preliminary study of transcriptional recovery kinetics after HS provides key insights into the field of stress response biology. Organisms naturally encounter various sources of stresses including HS, which they deal by activating a cellular defense mechanism known as HSR. Successful

recovery form HS develops an acquired tolerance to thermal stress, enabling cells to perform better when they re-encounter HS again. However, the precise mechanism and the factors involved in thermotolerance are unclear.

In this study, we provide one possible mechanism that could at least partially contribute if not fully explain the acquired thermotolerance. Although transcription of genes and enhancers returns to basal level, a class of genes is characterized by increased level of paused Pol II after recovery from HS. Interestingly, these genes exhibit slowest induction kinetics during HS. Pol II pausing at promoter-proximal region potentiates genes for rapid and synchronous induction upon various stimuli. Accordingly, it is plausible that these genes with increased pausing will rapidly and synchronously induce when they re-encounter HS. As many of these genes encode anti-apoptotic factors, the faster-than-intrinsic induction kinetics due to increased pausing may allow enough time for cells to make anti-apoptotic factors and overcome apoptosis. This ability of cells to combat apoptosis may have been a contributing factor in acquired thermotolerance.

To conclusively prove our hypothesis, a PRO-seq time course during second round of HS in stress-recovered cells showing faster induction kinetics of the anti-apoptotic genes is necessary. This observation would be critical in attributing thermotoleranceto increased Pol II pausing. It is important to emphasize the limitation of our study as we only probed transcriptional changes in our attempt to understand thermotolerance. We are aware that

post-transcriptional and translational control, and protein stability can modulate thermotolerance. Further studies are required to examine these possibilities.

In this study, we primarily used immortalized mouse embryonic fibroblasts that have been cultured in non-native environment. This perhaps is not the ideal system for studying thermotolerance. More importantly, the fluctuation in glucose and growth factors can trigger mild stress response. Addition of fresh tissue-culture media is known to induce *HSPs* in HSF1-independent manner (Mahat and Lis, 2016). This raises the possibility that cells in tissue-culture to already gain thermotolerance due to repeated activation of mild HSR. This could also explain the lack of transcriptional changes in annotated genes and enhancers after recovery from HS. A well-controlled experiment using an animal model that has not been stressed before could provide a more definitive answer to whether or not there is a transcriptional change at genes and enhancers.

REFERENCES

Adelman, K., and Lis, J.T. (2012). Promoter-proximal pausing of RNA polymerase II: emerging roles in metazoans. Nat. Rev. Genet. *13*, 720–731.

Baird, N.A., Douglas, P.M., Simic, M.S., Grant, A.R., Moresco, J.J., Wolff, S.C., Yates, J.R., Manning, G., and Dillin, A. (2014). HSF-1-mediated cytoskeletal integrity determines thermotolerance and life span. Science *346*, 360–363.

Core, L.J., Martins, A.L., Danko, C.G., Waters, C.T., Siepel, A., and Lis, J.T. (2014). Analysis of nascent RNA identifies a unified architecture of initiation regions at mammalian promoters and enhancers. Nat Genet *46*, 1311–1320.

Core, L.J., Waterfall, J.J., and Lis, J.T. (2008). Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. Science *322*, 1845–1848.

Djebali, S., Davis, C.A., Merkel, A., Dobin, A., Lassmann, T., Mortazavi, A., Tanzer, A., Lagarde, J., Lin, W., Schlesinger, F., et al. (2012). Landscape of transcription in human cells. Nature *489*, 101–108.

Duarte, F.M., Fuda, N.J., Mahat, D.B., Core, L.J., Guertin, M.J., and Lis, J.T. (2016). Transcription factors GAF and HSF act at distinct regulatory steps to modulate stress-induced gene activation. Genes & Development *30*, 1731–1746.

Kregel, K.C. (2002). Heat shock proteins: modifying factors in physiological stress responses and acquired thermotolerance. J. Appl. Physiol. *92*, 2177–2186.

Levine, M. (2011). Paused RNA Polymerase II as a Developmental Checkpoint. Cell *145*, 502–511.

Lindquist, S., and Kim, G. (1996). Heat-shock protein 104 expression is sufficient for thermotolerance in yeast. Proc. Natl. Acad. Sci. U.S.a. *93*, 5301–

5306.

Mahat, D.B., and Lis, J.T. (2016). Use of conditioned media is critical for studies of regulation in response to rapid heat shock. Cell Stress Chaperones 1–8.

Mahat, D.B., Salamanca, H.H., Duarte, F.M., Danko, C.G., and Lis, J.T. (2016). Mammalian Heat Shock Response and Mechanisms Underlying Its Genome-wide Transcriptional Regulation. Molecular Cell *62*, 63–78.

McMillan, D.R., Xiao, X., Shao, L., Graves, K., and Benjamin, I.J. (1998). Targeted disruption of heat shock transcription factor 1 abolishes thermotolerance and protection against heat-inducible apoptosis. J. Biol. Chem. *273*, 7523–7528.

Plank, J.L., and Dean, A. (2014). Enhancer Function: Mechanistic and Genome-Wide Insights Come Together. Molecular Cell *55*, 5–14.

Queitsch, C., Hong, S.W., Vierling, E., and Lindquist, S. (2000). Heat shock protein 101 plays a crucial role in thermotolerance in Arabidopsis. Plant Cell *12*, 479–492.

Seila, A.C., Calabrese, J.M., Levine, S.S., Yeo, G.W., Rahl, P.B., Flynn, R.A., Young, R.A., and Sharp, P.A. (2008). Divergent transcription from active promoters. Science *322*, 1849–1851.

Shui, W., Xiong, Y., Xiao, W., Qi, X., Zhang, Y., Lin, Y., Guo, Y., Zhang, Z., Wang, Q., and Ma, Y. (2015). Understanding the Mechanism of Thermotolerance Distinct From Heat Shock Response Through Proteomic Analysis of Industrial Strains of Saccharomyces cerevisiae. Mol. Cell Proteomics *14*, 1885–1897.

Singer, M.A., and Lindquist, S. (1998). Thermotolerance in Saccharomyces cerevisiae: the Yin and Yang of trehalose. Trends Biotechnol. *16*, 460–468.

Trent, J.D. (1996). A review of acquired thermotolerance, heat-shock proteins,

and molecular chaperones in archaea. FEMS Microbiology Reviews *18*, 249–258.

Wang, X., Lee, C., Gilmour, D.S., and Gergen, J.P. (2007). Transcription elongation controls cell fate specification in the Drosophila embryo. Genes & Development *21*, 1031–1036.

Zhang, Y., Huang, L., Zhang, J., Moskophidis, D., and Mivechi, N.F. (2002). Targeted disruption ofhsf1 leads to lack of thermotolerance and defines tissue-specific regulation for stress-inducible Hsp molecular chaperones. J. Cell. Biochem. *86*, 376–393.

CHAPTER 5[c]

USE OF CONDITIONED MEDIA IS CRITICAL FOR STUDIES OF REGULATION IN RESPONSE TO RAPID HEAT SHOCK

Summary

Heat shock response (HSR) maintains and restores protein homeostasis when cells are exposed to proteotoxic heat stress. Heat shock (HS) triggers a rapid and robust change in genome-wide transcription, protein synthesis, and chaperone activity; and therefore, the HSR has been widely used as a model system in these studies. The conventional method of performing instantaneous HS in the laboratory uses heated fresh media to induce HSR when added to cells. However, addition of fresh media to cells may evoke additional cellular responses and signaling pathways. Here, we compared the change in global transcription profile when HS is performed with either heated fresh media or heated conditioned media. We found that the use of heated fresh media induces transcription of hundreds of genes that HS alone does not induce, and masks or partially masks HS-mediated downregulation of thousands of genes. The fresh-media-dependent upregulated genes encode ribosomal subunit proteins involved in translation

[c] Adapted from Mahat, D. B., & Lis, J. T. (2016). Use of conditioned media is critical for studies of regulation in response to rapid heat shock. *Cell Stress & Chaperones*, 1–8. Reprinted with permission from Springer.

and RNA processing factors. More importantly, fresh media also induces transcription of several heat shock protein genes (*Hsps*) in a heat shock factor 1 (HSF1) independent manner. Thus, we conclude that a conventional method of HS with heated fresh media causes changes in transcription regulation that confound the actual change caused solely by elevated temperature of cells.

Introduction

Exposure of organisms to various kinds of cellular stresses is an inevitable consequence of living in a dynamic environment. Stresses disrupt protein homeostasis and give rise to a proteotoxic cellular environment, which is detrimental to the proper cellular functions (Morimoto, 1998). To cope with stresses such as elevated temperature, heavy metals, toxins, oxidants, and bacterial and viral infections, organisms mount a rapid and robust response, known as the heat shock response (HSR). This cellular defense mechanism against stresses is evolutionarily conserved in organisms of all complexities (Lindquist, 1986). The characteristic feature of the HSR is the dramatic induction of heat shock protein (HSP) genes (*Hsps*), which are molecular chaperones specialized in proper folding of *de novo* synthesized peptides and refolding of misfolded proteins (Morimoto, 1993). Successful rescue by *Hsps* to regain protein homeostasis returns cells to their native state; however, cells subjected to a prolonged proteotoxic environment can lead to a failure of

molecular chaperones to maintain proteostasis and eventually trigger cellular apoptosis pathways (Jolly and Morimoto, 2000).

In mammals, the rapid and robust induction of *Hsps* upon stress is regulated by heat shock transcription factor 1 (HSF1) (Wu, 1995) at the transcriptional level, which is a primary and major point of regulation (Lis and Wu, 1993). Upon stress, HSF1 trimerizes and binds to the promoter of *Hsps* at three or more inverted repeats of nGAAn pentamers (Perisic et al., 1989), known as the heat shock element (HSE). HSF1 recruits co-factors that promote release of paused Pol II into productive elongation (Lis et al., 2000). Increasing evidence suggests that the number of HSF1 regulated genes exceed the repertoire of genes referred to as *Hsps* (Trinklein et al., 2004; Vihervaara and Sistonen, 2014). HSF1 and its target genes are implicated in obesity (Ma et al., 2015), longevity (Morley and Morimoto, 2004), and cardiovascular diseases (Yan et al., 2002). More importantly, HSF1 plays opposing roles in cancer and neurological disorder (Ciocca and Calderwood, 2005; Dai et al., 2007). These findings support a broad role of HSF1 and HSR in homeostasis and disease. Nonetheless, much work remains in elucidating the totality of the stress response and its interplay with disease states.

The field of stress biology has greatly benefited from using the HSR as a model to study stress response. Presence and importance of the HSR from lower eukaryotes to mammals, similarity in factors involved, and the ease of use in laboratory settings makes it highly desirable system. HSR is robust,

extensive, and rapid: previous studies have documented recruitment of transcription factors and major transcriptional changes occurring within a few minutes of HS (O'Brien and Lis, 1993; Zobeck et al., 2010). Therefore, a precise method to instantly trigger HS in laboratory settings is important to further investigate the first-order kinetic changes in HS induced transcription regulation. To perform instantaneous HS, many labs substitute the existing cell-culture media (hereafter called conditioned media) with heated fresh media for adherent cells or add heated fresh media to cells in suspension. Although this strategy achieves instantaneous HS, it also possesses a risk, as addition of heated fresh media could evoke other cellular responses beyond the HSR. Transcriptional changes induced by treating serum-starved cells with fresh media are well documented (Pirkmajer and Chibalin, 2011), and substituting conditioned media with heated fresh media could induce a similar response. The potential unsolicited transcriptional changes caused by heated fresh media have not been explored.

We recently reported a robust and simple method of performing instantaneous HS using heated conditioned media from same cells growing for same period in same conditions (Mahat et al., 2016a), and we documented the changes in global distribution of transcriptionally engaged RNA polymerase at a base-pair resolution by our Precision Run-on and sequencing (PRO-seq) method (Kwak et al., 2013), In the present manuscript, we compare the use of conditioned media and fresh media in transcription profile changes during HS.

We performed PRO-seq in WT and *Hsf1*$^{-/-}$ mouse embryonic fibroblasts (MEFs) before and 12 minutes and 60 minutes after HS with either fresh media or conditioned media, both heated to 42^{0}C. We found that the transcription induction of some genes is entirely dependent on fresh media but not HS when fresh media is used for HS. We also discovered that the global change in transcription profile upon HS is severely confounded by the use of fresh media. More importantly, fresh media induces transcription of heat-inducible *Hsps* in both WT and *Hsf1*$^{-/-}$ MEFs indicating the activation of an HSF1-independent pathway for *Hsps* expression by some component of fresh media. Our work reveals potential pitfalls in experimental design of stress response studies in laboratory settings and strongly supports the use of heated conditioned media for instantaneous HS experiments.

Materials and methods

Cell culture

WT and *Hsf1*$^{-/-}$ MEFs generated from wild type mice and *Hsf1*$^{-/-}$ littermate were provided by Ivor Benjamin lab (McMillan et al., 1998). MEFs were grown in Dulbecco's Modified Eagle Medium supplemented with 10% heat inactivated fetal bovine serum (v/v) and 1% Penicillin Streptomycin (v/v) at 37^{0}C with 5% CO_2 and 90% humidity. Cells were grown to ~80% confluence for HS treatment in 150mm TC-treated and gamma irradiated cell culture dish.

Heat shock with fresh media

To harvest non-heat shocked cells, conditioned media from the cell plates were discarded and cells were rinsed with ice-cold PBS. To perform qPCR, cells were harvested by elution in 1 ml of Trizol. To prepare samples for PRO-seq, nuclei were isolated from the cells as described in the nuclei isolation section.

To perform instantaneous heat shock using fresh media, conditioned media from cells were discarded. 20 ml of fresh media heated to 42^0C was gently added to the cell culture plate and immediately placed in 42^0C incubator with 5% CO_2 and 90% humidity. The cells were heat shocked for desired time and harvested as described in the nuclei isolation section.

Heat shock with conditioned media

Non-heat shocked cells were harvested as described above. Conditioned media from the cell plates were discarded and cells were rinsed with ice-cold PBS. To perform qPCR, cells were harvested in 1 ml of Trizol. To prepare samples for PRO-seq, nuclei were isolated from the cells as described in the nuclei isolation section.

Heat shock with conditioned media was performed as described previously (Mahat et al., 2016a). It requires two sets of identical cell plates for a HS experiment. Conditioned media from the first cell plate was collected and heated to 42^0C in water bath. The cells on this cell plate are discarded. Once

the conditioned media was heated to 42^0C, conditioned media from the second cell plate was discarded and 20 ml of heated conditioned media collected from the first cell plate was gently added, and immediately placed in 42^0C incubator with 5% CO_2 and 90% humidity. The cells were heat shocked for desired time and harvested as described in the nuclei isolation section.

Nuclei isolation, nuclear run-on, and PRO-seq library preparation

Isolation of nuclei from WT and *Hsf1*$^{-/-}$ MEFs, nuclear run-on reactions, and PRO-seq library preparation were performed as described previously (Mahat et al., 2016b).

Differential expression analysis

DESeq2 was used to identify differentially expressed genes (Love et al., 2014). PRO-seq reads were calculated from the body of genes (+500 from Transcription start sites to -100 from polyA site) in both biological replicates and the differential expression was calculated between NHS and HS conditions. Genes with p-value less than 0.001 were considered to be significantly changed.

Gene ontology analysis

Differentially regulated genes in fresh media versus conditioned media were analyzed for enriched classes of molecular functions and biological

processes using the database for annotation, visualization and integrated discovery (DAVID) (Huang et al., 2009). Highly enriched and non-redundant GO terms from the PANTHER database were selected using the option provided in DAVID.

Results

Heated fresh media induces immediate early genes

Instantaneous heat shock is typically performed using heated fresh media (Figure 5.1A). When working with adherent cells (such as mammalian fibroblasts and Hela cells), conditioned media from the cells is discarded and fresh media heated to a desired temperature is added. Then the cells are placed in the heated incubator for a desired period of time (however, this practice is not always explicitly stated). When working with cells in suspension (such as Drosophila S2 cells and human K562 cells), adequate volume of heated fresh media is added to the existing media in order to attain the desired temperature (Kim and Gross, 2013; O'Brien and Lis, 1993; Zhang et al., 2014). These approaches of using heated fresh media for HS ensures the cells are instantaneously heat shocked; however, additional cellular responses and signaling pathways could be triggered by fresh serum or glucose present in fresh media and eventually confound the HSR. Therefore, to avoid the unwanted effects of fresh media for instantaneous HS, we used conditioned media for HS (Figure 5.1B). For each experiment, two sets of cells are plated.

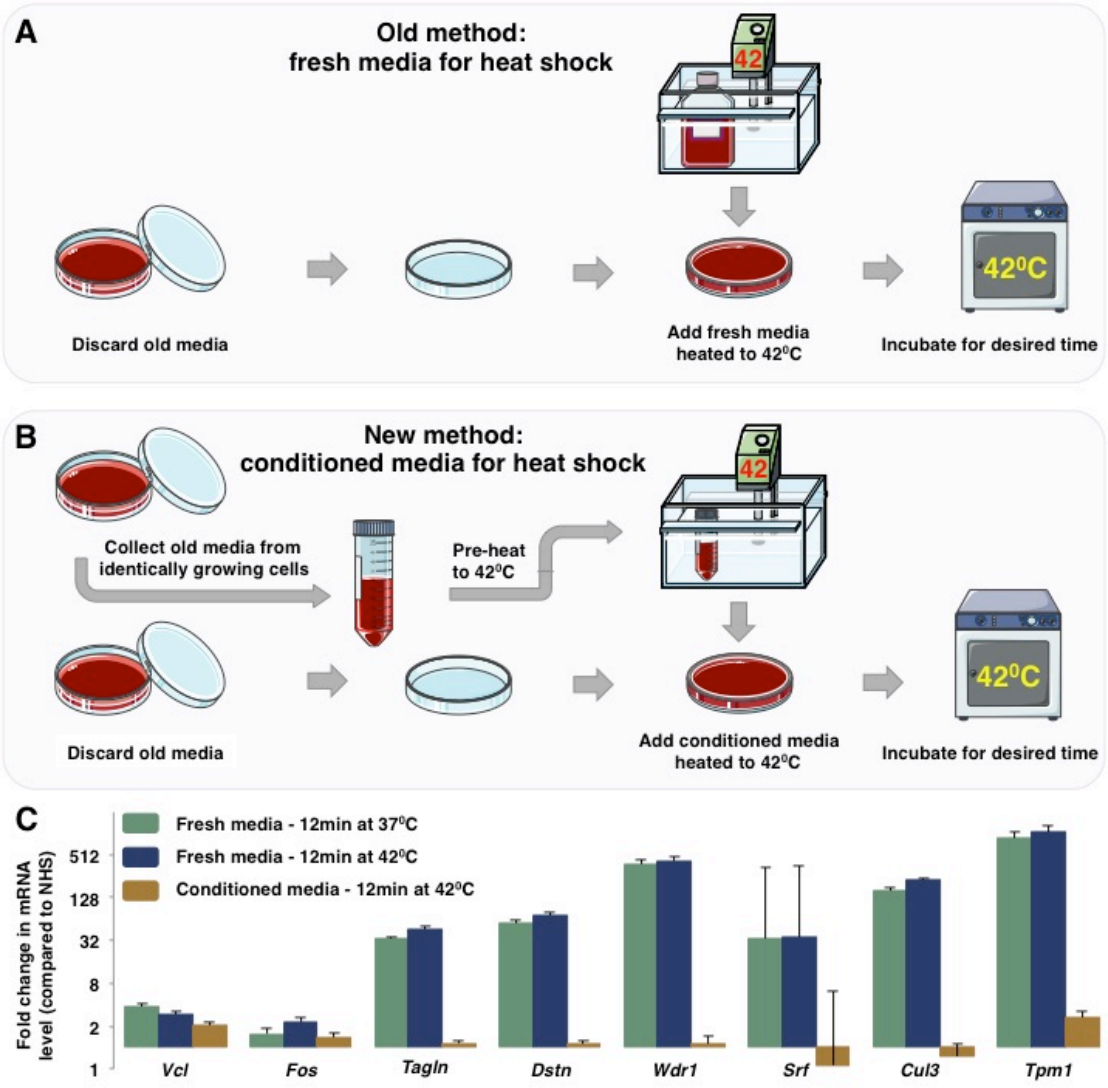


Figure 5.1. Fresh media transcriptionally induces IE genes
(A) Schematics of the conventional method of instantaneous HS. Conditioned media is discarded from cells and fresh media heated to the desired temperature is added to instantaneously HS the cells.
(B) Schematic of HS method using heated conditioned media. Conditioned media from one of a pair of identically growing cell cultures is collected and heated to the desired temperature. Conditioned media from the second set of cells is discarded and heated conditioned media is added to instantaneously HS the cells.
(C) Relative change in transcription of eight IE genes as measured by RT-qPCR in three conditions: i) fresh media for 12 min at 37^0C, ii) fresh media for 12 min at 42^0C and, iii) conditioned media for 12 min at 42^0C. Error bars represent standard deviation from three qPCR replicates.

Conditioned media from the first set of identically growing cells is collected and heated to a desired temperature. Once the media reaches the desired temperature, conditioned media from the second set of cells is discarded and heated conditioned media from the first set of cells is added to it, followed by incubation in a cell-culture incubator set to the desired temperature for the desired time. This approach ensures instantaneous HS and eliminates the possibility of unwanted transcriptional changes that could be caused by the use of fresh media.

To examine the effect of fresh media, we measured transcription change on a set of immediate early (IE) genes (Esnault et al., 2014). Addition of fresh media after serum-starvation immediately and transiently induces transcription of these genes (Schratt et al., 2001). To distinguish the contribution of fresh media from HS, we treated MEFs in three different ways: i) adding fresh media without HS, ii) adding fresh media and HS, and iii) adding conditioned media and HS, and then examined the changes in transcription level of eight IE genes compared to NHS cells (Figure 5.1C). We found that the transcription induction of these genes is dependent on fresh media, and HS has little to no effect on their transcription induction. This observation highlights the confounding effects caused by the use of fresh media for instantaneous HS in transcriptional induction of at least a few genes.

Fresh media causes significant change in genome-wide transcription profile

To examine the genome-wide transcriptional changes caused by the use of heated fresh media for HS, we performed PRO-seq using heated fresh media and compared with PRO-seq that was performed using heated conditioned media (Mahat et al., 2016a). We found that the fresh media used in HS strongly induces transcription of some IE genes, confirming the qPCR results. For example, transcription induction of *Nr4a1* gene is entirely dependent on fresh media, whereas transcription of *Vcl* gene is slightly induced by HS alone, which is further increased by fresh media (Figure 5.2A).

We measured the PRO-seq density change in all genes prior and after HS using DESeq2 (Love et al., 2014). An additional number of genes are upregulated at both 12 min and 60 min of HS when heated fresh media is used instead of heated conditioned media (Figure 5.2B). Moreover, ~25% of actively transcribing genes are downregulataed after 12 min of HS using fresh media; however, many of these genes are no longer downregulated at 60 min of HS indicating the transient nature of downregulation caused by the fresh media. In contrast, transcription downregulation during HS using conditioned media is gradual, and by 60 min of HS, ~55% of the actively transcribing genes are repressed.

To identify genes that are dramatically affected by the condition of media used for HS, we analyzed significantly upregulated genes at 12 min of

Figure 5.2. Genome-wide transcription profile change caused by fresh media is different than caused by HS

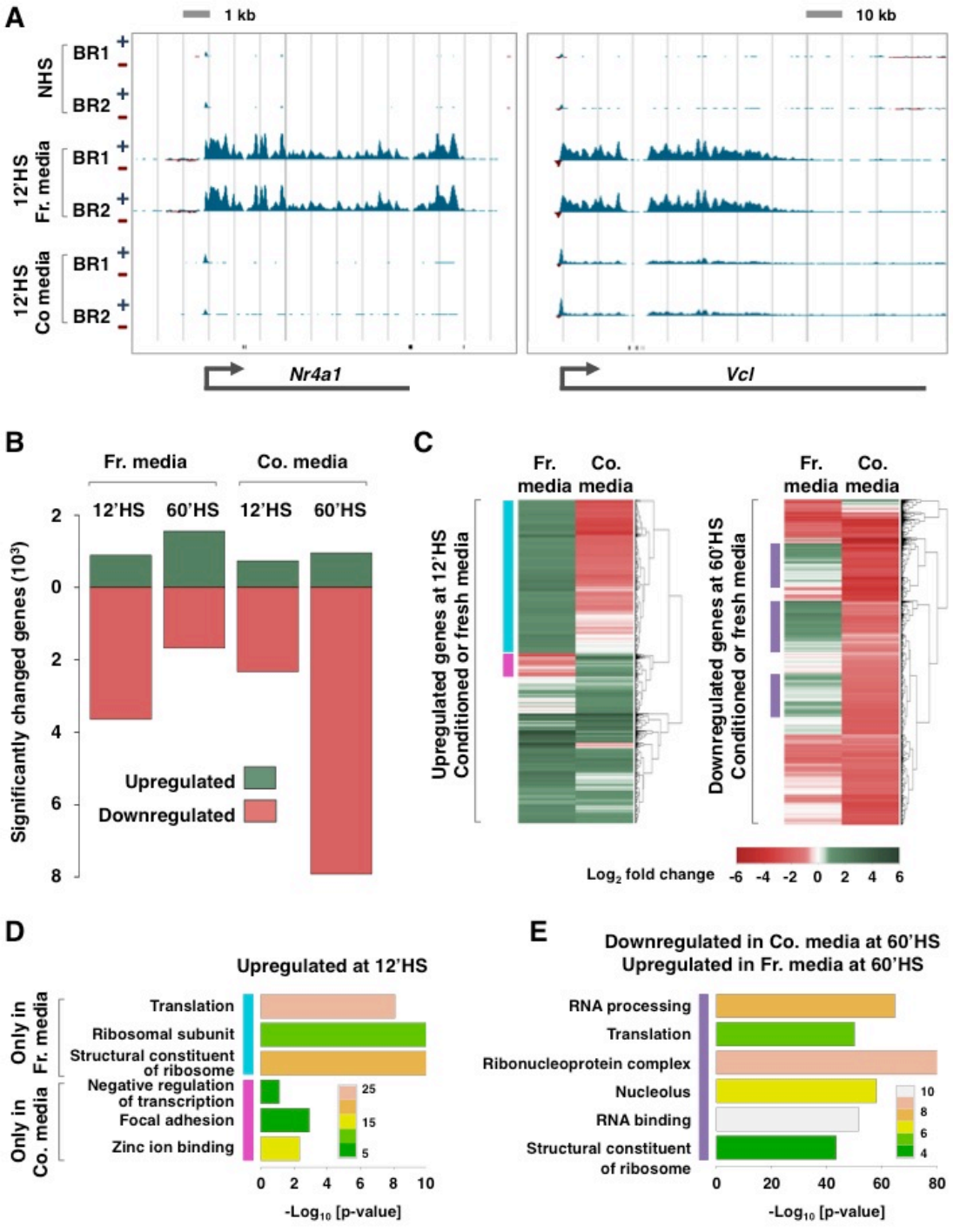
(A) Genome-browser screenshots of PRO-seq density in three conditions with two biological replicates each. PRO-seq density in sense and antisense direction is shown in blue and red respectively. Small vertical black bars at the bottom represent the genomic regions that do not map uniquely at 36 bp resolution.

(B) Number of significantly changed genes using heated fresh media or heated conditioned media for HS. Upregulated genes are shown in green and downregulated genes are shown in red.

(C) Heatmap showing the status of PRO-seq density change in all genes that show significant upregulation at 12 min of HS with either fresh or conditioned media (left) and all genes that show significant downregulation at 60 min of HS with either fresh or conditioned media (right). The color bars represent genes that were examined in gene ontology analysis in (d).

(D) Gene ontology analysis of enriched molecular functions and biological processes in significantly upregulated genes at 12 min of HS with fresh media but significantly downregulated in conditioned media is indicated by the blue bar, while significantly downregulated genes at 12 min of HS with fresh media but significantly upregulated in conditioned media are indicated by the magenta bar.

(E) Gene ontology analysis of enriched molecular functions and biological processes in significantly upregulated genes at 60 min of HS with fresh media but significantly downregulated in conditioned media (purple bar).


A
1 kb
10 kb
NHS
BR1
BR2
12'HS Fr. media
BR1
BR2
12'HS Co media
BR1
BR2
Nr4a1
Vcl
B
Fr. media
Co. media
12'HS
60'HS
12'HS
60'HS
Significantly changed genes (10³)
2
0
2
4
6
8
Upregulated
Downregulated
C
Fr. media
Co. media
Upregulated genes at 12'HS Conditioned or fresh media
Fr. media
Co. media
Downregulated genes at 60'HS Conditioned or fresh media
Log₂ fold change
-6 -4 -2 0 2 4 6
D
Upregulated at 12'HS
Only in Fr. media
Translation
Ribosomal subunit
Structural constituent of ribosome
Only in Co. media
Negative regulation of transcription
Focal adhesion
Zinc ion binding
25
15
5
0 2 4 6 8 10
-Log₁₀ [p-value]
E
Downregulated in Co. media at 60'HS
Upregulated in Fr. media at 60'HS
RNA processing
Translation
Ribonucleoprotein complex
Nucleolus
RNA binding
Structural constituent of ribosome
10
8
6
4
0 20 40 60 80
-Log₁₀ [p-value]

HS and significantly downregulated genes at 60 min of HS in either fresh or conditioned media (Figure 5.2C). This side-by-side visualization of regulated genes in two different media conditions allows characterization of genes that undergo complete reversal in regulation. While some genes upregulated at 12 min of HS in conditioned media are also upregulated in fresh media, many genes downregulated in conditioned media are upregulated in fresh media (Figure 5.2C, left panel - blue bar), and a few genes upregulated in conditioned media are downregulated in fresh media (Figure 5.2C, left panel - magenta bar). Similarly, some genes that are significantly downregulated at 60 min of HS in conditioned media are upregulated in fresh media (Figure 5.2C, right panel - purple bar). These three classes, marked by magenta, blue, and purple bar, represent genes whose activities are either induced or repressed depending on the external cues.

Intrigued by this dramatic change in transcriptional regulation that is dependent on the media used for HS, we examined the above three classes of differentially regulated genes for their cellular functions using gene ontology (GO) analysis. The upregulated genes only in fresh media at 12 min of HS (blue bar in Figure 5.2C) are enriched for ribosomal subunits, its structural components, and translation (Figure 5.2D). In contrast, upregulated genes only in conditioned media at 12 min of HS (magenta bar in Figure 5.2C) are enriched for zinc ion binding proteins, negative regulators of transcription, and focal adhesion genes. Similarly, downregulated genes in conditioned media at

60 min of HS but induced in fresh media (purple bars in Figure 5.2C) are enriched for RNA binding, RNA processing, and ribosomal structures involved in translation machinery (Figure 5.2E). While these findings conform the general consensus in the field of stress biology that global transcription and translation are repressed during HS (Anckar and Sistonen, 2011), it also indicates that the fresh media overrides the HS-mediated repression and in turn activates these genes involved in increase mRNA and protein production. Overall, this observation further supports our hypothesis that heated conditioned media should be used for instantaneous HS to avoid the confounding regulation caused by fresh media.

Fresh media induces Hsps in an HSF1-independent manner

The broad class of *Hsps* includes 75 genes that play a role in protein homeostasis (Kampinga et al., 2009). To dissect the role of fresh media in transcription regulation of *Hsps*, we examined the PRO-seq density change in all 75 genes upon HS with heated fresh media or heated conditioned media. In order to access the role of HSF1 in transcription regulation of *Hsps*, we also performed PRO-seq in Hsf1 knockout (*Hsf1*$^{-/-}$) MEFs. We focused our analysis on 48 genes that are induced in at least one time-point or cell type during HS. ~33% of these genes, including major *Hsps*, are robustly induced in WT MEFs while almost all of the 48 genes are either downregulated or unchanged in *Hsf1*$^{-/-}$ MEFs in conditioned media (Figure 5.3A). In contrast, most of these 48

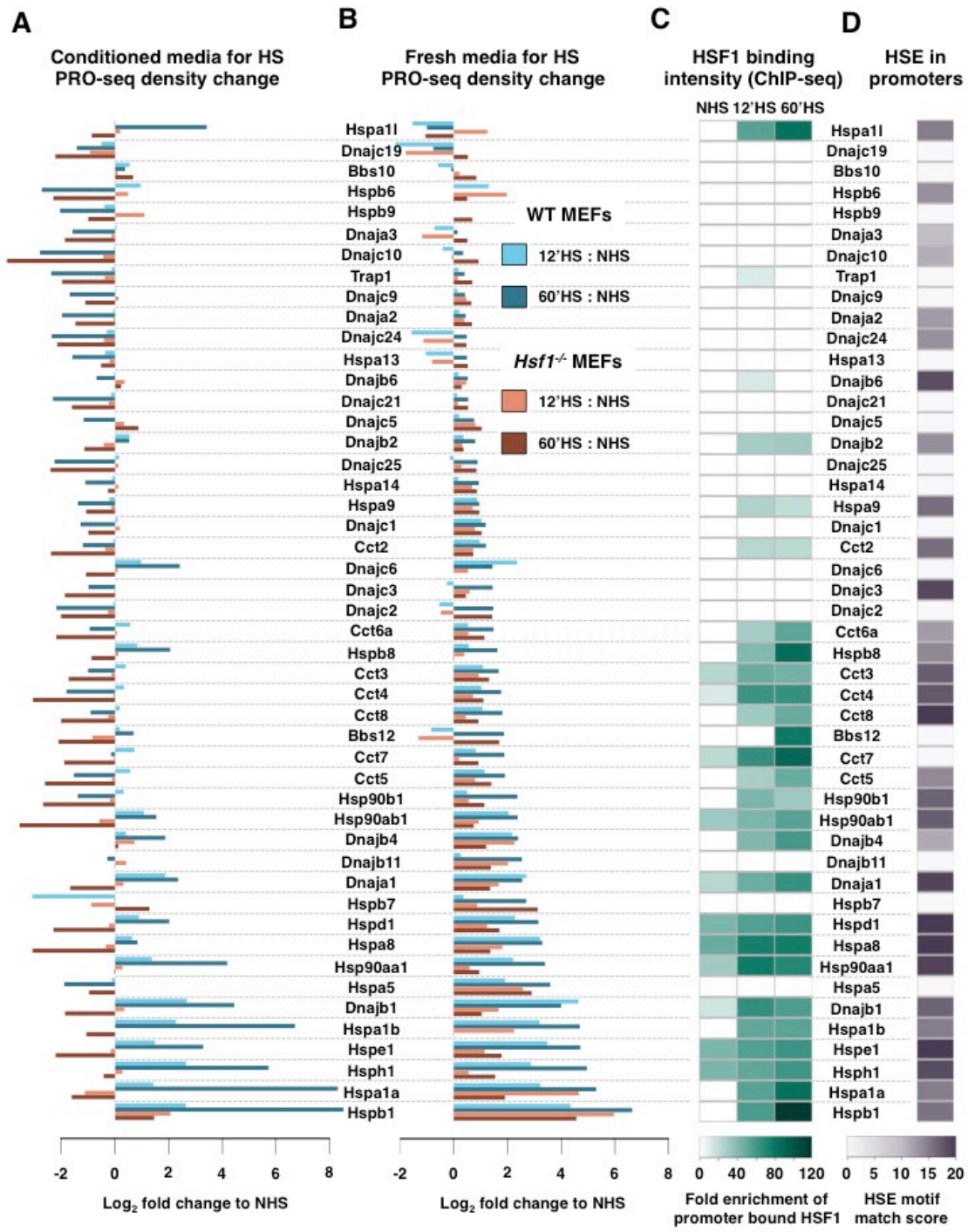
Figure 5.3. Fresh media induces many *Hsps* in HSF1-independent manner

(A) PRO-seq density change during HS with conditioned media in 48 genes within the broad HSP family in WT and *Hsf1*$^{-/-}$ MEFs.

(B) PRO-seq density change during HS with fresh media in the corresponding 48 genes in WT and *Hsf1*$^{-/-}$ MEFs.

(C) HSF1 fold enrichment, as measured by HSF1 ChIP-seq, in the promoter of the corresponding 48 genes in WT MEFs.

(D) HSE motif match score in the promoters of the corresponding 48 genes.


A
Conditioned media for HS
PRO-seq density change
B
Fresh media for HS
PRO-seq density change
C
HSF1 binding
intensity (ChIP-seq)
D
HSE in
promoters
NHS 12'HS 60'HS
WT MEFs
12'HS : NHS
60'HS : NHS
Hsf1-/- MEFs
12'HS : NHS
60'HS : NHS
Hspa1l
Dnajc19
Bbs10
Hspb6
Hspb9
Dnaja3
Dnajc10
Trap1
Dnajc9
Dnaja2
Dnajc24
Hspa13
Dnajb6
Dnajc21
Dnajc5
Dnajb2
Dnajc25
Hspa14
Hspa9
Dnajc1
Cct2
Dnajc6
Dnajc3
Dnajc2
Cct6a
Hspb8
Cct3
Cct4
Cct8
Bbs12
Cct7
Cct5
Hsp90b1
Hsp90ab1
Dnajb4
Dnajb11
Dnaja1
Hspb7
Hspd1
Hspa8
Hsp90aa1
Hspa5
Dnajb1
Hspa1b
Hspe1
Hsph1
Hspa1a
Hspb1
-2 0 2 4 6 8
Log2 fold change to NHS
-2 0 2 4 6 8
Log2 fold change to NHS
0 40 80 120
Fold enrichment of
promoter bound HSF1
0 5 10 15 20
HSE motif
match score

genes are transcriptionally induced during HS with heated fresh media (Figure 5.3B). More importantly, transcription of these genes is also induced in *Hsf1*$^{-/-}$ MEFs. The profile of HSF1 binding to the promoter of the genes (Figure 5.3C) and the presence of HSE in the corresponding promoters (Figure 5.3D) resembled the induced *Hsps* with conditioned media, which further indicates that the induction of *Hsps* with fresh media is mediated by factors other than HSF1. Together, these findings reveal the potential of some component of fresh media to transcriptionally induce *Hsps* in HSF1-independent manner that are either repressed or unchanged during a proper, conditioned-media HS.

Discussion

HS is widely used as the model system to study stress response. It triggers a rapid and robust transcriptional response, and it is convenient to use in a laboratory settings. However, the widespread approach of using heated fresh media for instantaneous HS carries a risk of undesired transcriptional changes. Recently, we demonstrated a simple but effective method of performing instantaneous heat shock using conditioned media from same cells growing in identical conditions (Mahat et al., 2016a) . Here, we now report the changes in transcription profile during HS when heated fresh media or heated conditioned media is used for instantaneous HS. In genome-wide transcriptional analysis using PRO-seq, we found that fresh media causes HS-independent transcription induction of additional genes and also mediates HS-

independent transient repression of several genes. These fresh media dependent genes are enriched in ribosomal machinery components and mRNA translation. More importantly, we found that fresh media induces many *Hsps* that are either repressed or unchanged transcriptionally during HS and this induction potential of fresh media is independent of HSF1. We also discovered that the heated fresh media used for instantaneous HS strongly induces some IE genes in addition to *Hsps.*

One defining feature of IE genes is the presence of serum response element (SRE) in their promoters, and all eight IE genes examined in this study contains strong SRE. In serum-stimulation studies, serum response factor (SRF) binds to SRE and transiently induces expression of IE genes. Therefore, it is possible that either fresh serum or fresh glucose in fresh media activates SRF, which is responsible for the dramatic induction of IE genes in this study. Nevertheless, some of the eight IE genes also showed small but detectable induction by HS alone. Because the common feature of these genes is the presence of SRE in their promoters, the HS dependent mild induction of these genes is also likely mediated by SRF. In fact, we recently identified SRF as a novel transcription factor activated by HS, which transiently binds to the promoters of transiently induced genes during HS (Mahat et al., 2016a). Based on these findings, we believe that HS itself is also capable of activating SRF. However, the mechanism of SRF activation by fresh media and HS could be different. Addition of fresh media after serum-starvation

activates mitogen-activated protein kinase (MAPK) pathway and SRF is a known target of MAPK pathway (Heidenreich et al., 1999). Whereas, the interaction of SRF with the nuclear imported MAL, which is released from monomeric actin (Miralles et al., 2003), possibly due to actin polymerization during HS, is the proposed mechanism of SRF activation during HS (Mahat et al., 2016a).

Fresh media transcriptionally induces another class of genes, which encode components of ribosomal machinery involved in mRNA translation. This could be an indication of likely increased demand in translation caused by the activation of cell proliferation signaling pathways by fresh media. Fresh media also induces some *Hsps* in an HSF1-independent manner, whereas HS only induces the major *Hsps*. One likely candidate for the broad induction of *Hsps* by fresh media is the activator protein 1 (AP-1) transcription factors (TFs). AP-1 TFs are activated by serum stimulation (Poimenidi et al., 2009), and AP-1 binding element is highly enriched in the proximity of HSEs present in the promoters of *Hsps* (Mahat et al., 2016a). The highly significant enrichment of AP-1 motifs around HSEs suggests either collaboration or redundancy between AP-1 and HSF1 in induction of *Hsps*. When fresh media is added, perhaps cells trigger HSF1-independent induction of *Hsps* through AP-1 in order to chaperone the influx of newly synthesized proteins to support cell proliferation.

The unwanted transcriptional changes caused by fresh media after instantaneous HS has eluded detection until now for several reasons. First, the major HSF1-dependent *Hsps*, which are the primary focus of HS studies, are induced to similar level when either fresh media or conditioned media is used for HS. Second, genome-wide assays such as RNA-seq and mass-spectrometry have contributions from post-transcriptional regulatory steps of mRNA stability and translation efficiency respectively and may not reflect the precise change in transcription level. Third, genomic tools with sensitivity required to detect small changes in nascent transcription were unavailable until now. In the last few years, highly sensitive genomic methods to quantify nascent transcription have been developed. These methods, such as GRO-seq (Core et al., 2008), PRO-seq (Kwak et al., 2013), and NET-seq (Churchman and Weissman, 2011), will help to investigate the outstanding questions in the field of stress response. Because of the sensitivity of these assays, careful consideration should be taken in experimental setups, and at the very least, use of heated fresh media for instantaneous HS should be avoided.

REFERENCES

Anckar, J., and Sistonen, L. (2011). Regulation of HSF1 function in the heat stress response: implications in aging and disease. Annu. Rev. Biochem. *80*, 1089–1115.

Churchman, L.S., and Weissman, J.S. (2011). Nascent transcript sequencing visualizes transcription at nucleotide resolution. Nature *469*, 368–373.

Ciocca, D.R., and Calderwood, S.K. (2005). Heat shock proteins in cancer: diagnostic, prognostic, predictive, and treatment implications. Cell Stress Chaperones *10*, 86–103.

Core, L.J., Waterfall, J.J., and Lis, J.T. (2008). Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. Science *322*, 1845–1848.

Dai, C., Whitesell, L., Rogers, A.B., and Lindquist, S. (2007). Heat shock factor 1 is a powerful multifaceted modifier of carcinogenesis. Cell *130*, 1005–1018.

Esnault, C., Stewart, A., Gualdrini, F., East, P., Horswell, S., Matthews, N., and Treisman, R. (2014). Rho-actin signaling to the MRTF coactivators dominates the immediate transcriptional response to serum in fibroblasts. Genes & Development *28*, 943–958.

Heidenreich, O., Neininger, A., Schratt, G., Zinck, R., Cahill, M.A., Engel, K., Kotlyarov, A., Kraft, R., Kostka, S., Gaestel, M., et al. (1999). MAPKAP kinase 2 phosphorylates serum response factor in vitro and in vivo. J. Biol. Chem. *274*, 14434–14443.

Huang, D.W., Sherman, B.T., and Lempicki, R.A. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc *4*, 44–57.

Jolly, C., and Morimoto, R.I. (2000). Role of the heat shock response and molecular chaperones in oncogenesis and cell death. J. Natl. Cancer Inst. *92*,

1564–1572.

Kampinga, H.H., Hageman, J., Vos, M.J., Kubota, H., Tanguay, R.M., Bruford, E.A., Cheetham, M.E., Chen, B., and Hightower, L.E. (2009). Guidelines for the nomenclature of the human heat shock proteins. Cell Stress Chaperones *14*, 105–111.

Kim, S., and Gross, D.S. (2013). Mediator recruitment to heat shock genes requires dual Hsf1 activation domains and mediator tail subunits Med15 and Med16. Journal of Biological Chemistry *288*, 12197–12213.

Kwak, H., Fuda, N.J., Core, L.J., and Lis, J.T. (2013). Precise maps of RNA polymerase reveal how promoters direct initiation and pausing. Science *339*, 950–953.

Lindquist, S. (1986). The heat-shock response. Annu. Rev. Biochem. *55*, 1151–1191.

Lis, J., and Wu, C. (1993). Protein traffic on the heat shock promoter: parking, stalling, and trucking along. Cell *74*, 1–4.

Lis, J.T., Mason, P., Peng, J., Price, D.H., and Werner, J. (2000). P-TEFb kinase recruitment and function at heat shock loci. Genes & Development *14*, 792–803.

Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. *15*, 550.

Ma, X., Xu, L., Alberobello, A.T., Gavrilova, O., Bagattin, A., Skarulis, M., Liu, J., Finkel, T., and Mueller, E. (2015). Celastrol Protects against Obesity and Metabolic Dysfunction through Activation of a HSF1-PGC1α Transcriptional Axis. Cell Metab.

Mahat, D.B., Salamanca, H.H., Duarte, F.M., Danko, C.G., and Lis, J.T. (2016a). Mammalian Heat Shock Response and Mechanisms Underlying Its Genome-wide Transcriptional Regulation. Molecular Cell *62*, 63–78.

Mahat, D.B., Kwak, H., Booth, G.T., Jonkers, I.H., Danko, C.G., Patel, R.K., Waters, C.T., Munson, K., Core, L.J., and Lis, J.T. (2016b). Base-pair-resolution genome-wide mapping of active RNA polymerases using precision nuclear run-on (PRO-seq). Nat Protoc *11*, 1455–1476.

McMillan, D.R., Xiao, X., Shao, L., Graves, K., and Benjamin, I.J. (1998). Targeted disruption of heat shock transcription factor 1 abolishes thermotolerance and protection against heat-inducible apoptosis. J. Biol. Chem. *273*, 7523–7528.

Miralles, F., Posern, G., Zaromytidou, A.-I., and Treisman, R. (2003). Actin dynamics control SRF activity by regulation of its coactivator MAL. Cell *113*, 329–342.

Morimoto, R.I. (1993). Cells in stress: transcriptional activation of heat shock genes. Science *259*, 1409–1410.

Morimoto, R.I. (1998). Regulation of the heat shock transcriptional response: cross talk between a family of heat shock factors, molecular chaperones, and negative regulators. Genes & Development *12*, 3788–3796.

Morley, J.F., and Morimoto, R.I. (2004). Regulation of longevity in Caenorhabditis elegans by heat shock factor and molecular chaperones. Mol. Biol. Cell *15*, 657–664.

O'Brien, T., and Lis, J.T. (1993). Rapid changes in Drosophila transcription after an instantaneous heat shock. Mol. Cell. Biol. *13*, 3456–3463.

Perisic, O., Xiao, H., and Lis, J.T. (1989). Stable binding of Drosophila heat shock factor to head-to-head and tail-to-tail repeats of a conserved 5 bp recognition unit. Cell *59*, 797–806.

Pirkmajer, S., and Chibalin, A.V. (2011). Serum starvation: caveat emptor. Am. J. Physiol., Cell Physiol. *301*, C272–C279.

Poimenidi, E., Hatziapostolou, M., and Papadimitriou, E. (2009). Serum stimulates Pleiotrophin gene expression in an AP-1-dependent manner in human endothelial and glioblastoma cells. Anticancer Res. *29*, 349–354.

Schratt, G., Weinhold, B., Lundberg, A.S., Schuck, S., Berger, J., Schwarz, H., Weinberg, R.A., Rüther, U., and Nordheim, A. (2001). Serum response factor is required for immediate-early gene activation yet is dispensable for proliferation of embryonic stem cells. Mol. Cell. Biol. *21*, 2933–2943.

Trinklein, N.D., Murray, J.I., Hartman, S.J., Botstein, D., and Myers, R.M. (2004). The role of heat shock transcription factor 1 in the genome-wide regulation of the mammalian heat shock response. Mol. Biol. Cell *15*, 1254–1261.

Vihervaara, A., and Sistonen, L. (2014). HSF1 at a glance. J. Cell. Sci. *127*, 261–266.

Wu, C. (1995). Heat shock transcription factors: structure and regulation. Annu. Rev. Cell Dev. Biol. *11*, 441–469.

Yan, L.J., Christians, E.S., Liu, L., Xiao, X.Z., Sohal, R.S., and Benjamin, I.J. (2002). Mouse heat shock transcription factor 1 deficiency alters cardiac redox homeostasis and increases mitochondrial oxidative damage. Embo J. *21*, 5164–5172.

Zhang, H., Gao, L., Anandhakumar, J., and Gross, D.S. (2014). Uncoupling transcription from covalent histone modification. PLoS Genet. *10*, e1004202.

Zobeck, K.L., Buckley, M.S., Zipfel, W.R., and Lis, J.T. (2010). Recruitment timing and dynamics of transcription factors at the Hsp70 loci in living cells. Molecular Cell *40*, 965–975.

CHAPTER 6

CONCLUSIONS AND FUTURE DIRECTIONS

Old questions, new tools, and novel insights

An organism's primal purpose, from evolutionary perspective, is to reproduce and survive. The primary factors that dictate survival are food availability, predator-prey relationship, and environmental changes. Long-term environmental variation provides an anatomical and physiological canvas for evolutionary forces to fiddle with broad strokes, whereas temporary fluctuation in ambient condition stimulates evolution of adaptive mechanisms. One such adaptive mechanism for protection against elevated temperature, known as heat shock response (HSR), is highly conserved across domains of life. Despite significant variation in optimal growth temperature among organisms, increase of few degree Celsius triggers HSR. Such precise calibration of ideal temperature for growth and survival underscores the delicate balance in protein structure and homeostasis. No wonder cells under stress engage entire suite of gene regulatory switches from transcription to translation to orchestrate a finely tuned gene expression program that boosts survival and unplugs non-critical activity.

Among the major regulatory switches administered during HSR, transcription is the earliest and the primary mode of regulation. Cells sense

heat stress through multiple non-overlapping mechanisms and ultimately activate heat shock transcription factor (HSF1 in vertebrates). Activated HSF1 trimerizes via a coiled-coiled central domain, gains competency to bind heat shock element (HSE – an inverted repeat of three nGAAn pentamers) in the promoters of heat shock protein genes (*Hsps*), and dramatically induces their expression. These molecular events in a nutshell represent transcriptional control of HSR, which were discovered over the span of four decades, largely through focused gene studies in lower eukaryotes.

The dawn of affordable sequencing spurred genome-wide analyses of gene expression changes during HSR. mRNA sequencing in cells undergoing heat shock (HS) revealed changes in steady-state mRNA pool, which represent cumulative effect of transcriptional and post-transcriptional regulation. It was only after the development of highly sensitive genomic assays, which measured transcription by directly quantifying actively transcribing RNA polymerases (Core et al., 2008; Kwak et al., 2013), that precise and comprehensive assessment of transcriptional changes during HS was made possible.

In work presented in this dissertation, we used precision nuclear run-on sequencing (PRO-seq) to measure transcriptional changes immediately and after full-blown HS in mouse embryonic fibroblasts that lacked either HSF1, HSF2, both, or none. Genome-wide data on transcriptional changes from PRO-seq was complemented with transcription factor binding information from

ChIP-seq. The major findings tabulated below are reported in detail in this dissertation (Chapters 2 & 3).

- HSR triggers rapid, large-scale, and diverse changes in transcription.
- Majority of the genes regulated upon HSR are independent of HSF1.
- Heat-inducible *HSP* genes are entirely dependent on HSF1 for transcription induction.
- HSF1 regulates transcriptional induction by facilitating the release of promoter-proximal paused Pol II.
- HSF1 does not contribute to the genome-wide repression of transcription.
- Genome-wide transcriptional repression during HSR is caused by the inhibition of paused Pol II release into productive elongation.
- Pause release factor P-TEFb occupancy increases at the promoters of HSF1-induced genes and decreases at the promoters of repressed genes.
- A class of anti-apoptotic genes is induced late during HSR.
- Cytoskeleton genes are activated as early as classical *HSP* genes.
- Serum response factor (SRF) transiently binds to the promoters of cytoskeletal gens and transiently induces their transcription.
- HSF2 does not compensate the loss of HSF1.
- ~20% of the actively transcribed genes are repressed in HSF2-dependent manner during HSR.
- HSF2 requires HSF1 for DNA binding.

The evaluation of HSF1's contribution during HSR indicated that HSF1 is more like a 'specialized paramedic' rather than the all-powerful 'fire chief'. While these new insights may have deflated HSF1's perceived global contribution in stress repsonse, they have undoubtedly unveiled the enormity and complexity of the HSR. Yet, many questions remain unanswered. What stimulates and what impairs promoter-bound HSF1's ability to induce transcription? Is the correlation between P-TEFb occupancy and transcriptional changes during HSR a causative relationship? Can P-TEFb explain both induction and repression of transcription? Are the genes without promoter-bound HSF1 but induced in HSF1-dependent manner regulated by enhancer bound HSF1? What other transcription factors besides HSF1 and SRF regulate transcription during HSR? Is the enrichment of Nuclear Factor Erythroid 2 Like 2 (NFE2L2) in the promoters of late-induce anti-apoptotic genes functional? These are some of the few unanswered questions for future pursuit that arose directly from the work presented on this dissertation, one of which is elaborated below.

Understanding P-TEFb's role in HSR

One particular preliminary finding that has a broad implication in regulation of the HSR is the role of P-TEFb. P-TEFb occupancy increases in HSF1-induced genes and decreases in repressed genes, while transcription of the P-TEFb subunit genes remain unchanged during HSR (unpublished data). This apparent redistribution of the limited amount of cellular P-TEFb prompts

us to speculate that the concerted recruitment of P-TEFb to the HS-induced genes by HSF1 and other transcription factors from the genes that will eventually end up being repressed due to P-TEFb's unavailability is the primary mechanism of transcription regulation during HSR. The two key experiments to test this hypothesis are listed below.

P-TEFb overexpression

If lack of P-TEFb availability is the underlying mechanism of global downregulation during HSR, overexpression of P-TEFb should rescue or mitigate the global repression. Overexpression of kinase-mutant P-TEFb would decipher whether P-TEFb's role in HSR is dependent on its kinase activity or mediated through kinase-independent domain-specific interactions.

Inhibition of HSF1's ability to recruit P-TEFb

P-TEFb is recruited to heat shock loci in HSF1-dependent manner (Lis et al., 2000). Inhibiting HSF1's ability to recruit P-TEFb by a synthetic compound KRIBB11 (Yoon et al., 2011) should attenuate transcription induction of HSF1-bound and dependent genes upon HS. Concurrently, because the HSF1 cannot recruit P-TEFb, the P-TEFb level should remain unaffected at other genes and the global repression should be mitigated simultaneously.

If this 'redistribution of limited P-TEFb' hypothesis can't be proved, other hypotheses of P-TEFb-mediated global regulation in HSR should be tested. The 'reset' model (Chapter 1), which argues an instantaneous

genome-wide dislodgement of engaged P-TEFb and the subsequent requirement of assisted delivery to selected genes for transcription induction by factors such as HSF1, can be tested by quantifying chromatin bound vs unbound P-TEFb levels immediately upon HS. P-TEFb could also be limiting either by reduction in its kinase activity, which is required for Pol II pause release, or due to its depletion from nucleus and sequestration in cytoplasmic stress granules.

A peek into human health through the window of HSR

The mechanisms and machinery of stress response, primarily evolved to protect against proteotoxic stress, have major implications in health and diseases. Research recommendations in three areas that deeply interest me and with direct relevance to human health are briefly explained below.

Thermotolerance

On Chapter 4 of this dissertation, I presented preliminary work on exploration of the underlying mechanism of thermotolerance. Do cells retain transcriptional memory upon recovery from stress? In addressing this physiologically relevant question, I discovered that the promoter-proximal Pol II pausing in anti-apoptotic genes increases upon recovery from stress, despite lack of change in gene-body and enhancer transcription globally. Interestingly, these anti-apoptotic genes are induced late during heat shock. These two observations put together suggest a possible mechanism of

thermotolerance in stress-recovered cells. Perhaps thermotolerance is manifested through rapid induction of anti-apoptotic genes that confer survival advantages. This hypothesis can be tested by the experiment below.

Using HS-recovered cells, examine the induction kinetics of anti-apoptotic genes during subsequent HS

Cells that have recovered from 1 hour of HS for at least 48 hours are re-exposed again to 1 hour of HS (Figure 6.1A). Cells are harvested at various time points during both rounds of HS (Figures 6.1B & 6.1C) to measure the kinetics of transcription induction. Induction kinetics of the anti-apoptotic genes are monitored by measuring PRO-seq wave using a three-state hidden Markov model (HMM) (Danko et al., 2013).

If the anti-apoptotic genes in HS-recovered cells indeed exhibit faster than their normal induction kinetics in a subsequent HS, the finding would be a first step in establishing the functional consequence of transcriptional memory in the form of paused Pol II in thermotolerance. The second step would be to test whether the rapid induction of the anti-apoptotic genes confers survival advantage in HS-recovered cells undergoing subsequent HS.

In case of the disproval of this hypothesis, other likely explanations such as change in chromatin marks should be examined. Because pausing is a manifestation of promoter architecture, increased pausing in selected genes upon recovery from stress could be a result of change in chromatin structure and histone marks.

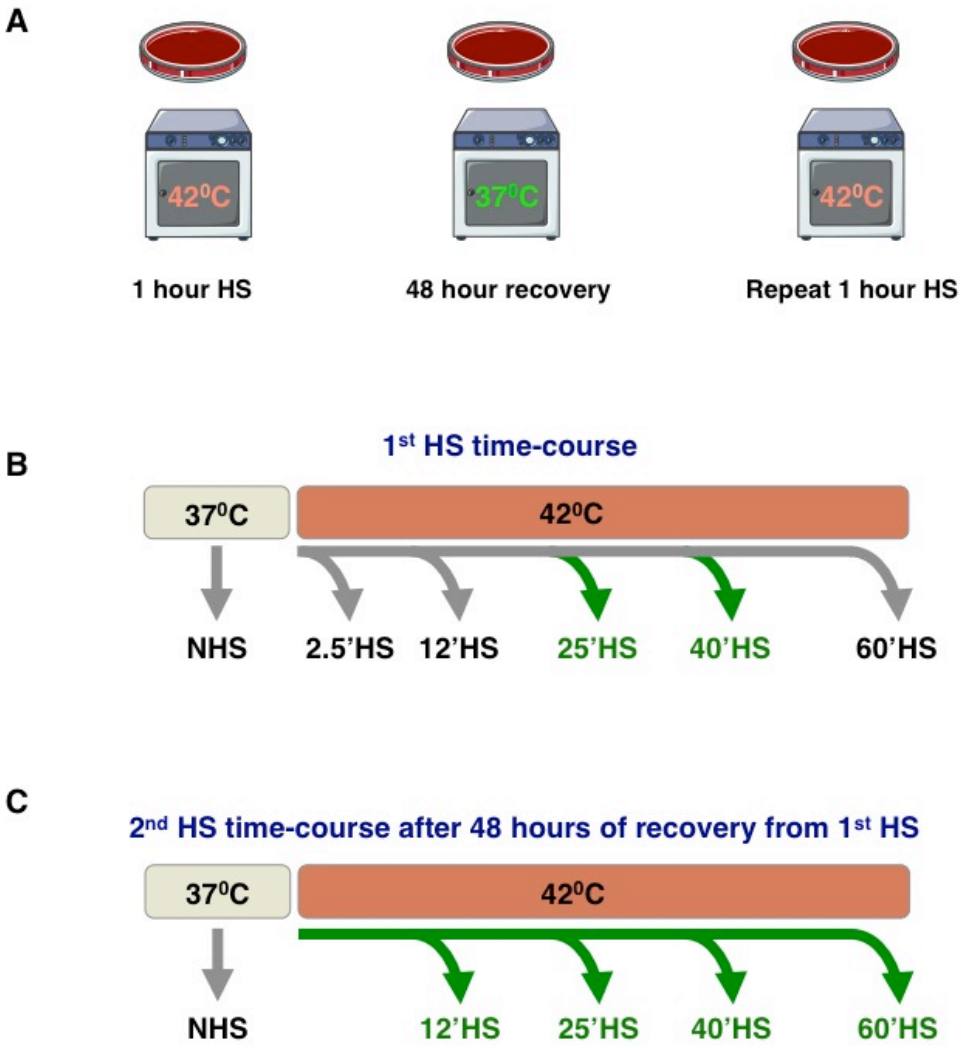


Figure 6.1. Design of thermotolerance experiment
(A) Schematics of re-exposing HS-recovered cells to new HS.
(B) HS time course performed on first HS. PRO-seq data needs to be collected for the time points shown in green. PRO-seq data exists for time points shown in grey.
(C) HS time course for new HS of HS-recovered cells.

Cancer

Cancer cells are characterized by higher level of HSF1 in tumor samples (Santagata et al., 2011). However, little is known about the molecular switches that drive increased expression of HSF1 in cancer. Conversely, the effects of over-expression or constitutive-activation of HSF1 in promoting cellular transformation remain to be explored. Whether high level of HSF1 in highly malignant cells is the cause or the consequence of transformation has a profound significance in design of HSF1-based cancer therapeutics.

Role of increased HSF1 level or activity in transformation

HSF1 is over-expressed by introducing ectopic copy or constitutively-activated through substitution of hydrophobic residues in heptad repeats (Rabindran et al., 1993; Zuo et al., 1994). Transformation cues below the transformation threshold are provided and transformation efficiency in presence and absence of HSF1 over-expression or constitutive-activation is compared. Increased transformation efficiency in cells with HSF1 over-expression or constitutive-activation would support our hypothesis.

This study would provide insights on whether some cellular environment can be more conducive for transformation due to activating mutations on HSF1 or copy number amplification of HSF1.

Transcription read-through

Inability of RNA polymerase to initiate or to terminate transcription at predefined sites results in generation of non-canonical transcripts. The

transcription read-through beyond the termination site generates fusion transcripts in breast, prostate, gastric, and kidney cancers (Choi et al., 2016; Grosso et al., 2015; Kim et al., 2015; Nacu et al., 2011). The RNA chimeras formed by transcription invasion of neighboring genes can encode novel oncogenic proteins (Valentijn et al., 2006), and the higher degree of transcription read-through correlates with poorer prognosis in renal carcinoma patients (Grosso et al., 2015).

Transcription read-through is also prominent in HSR. Transcription of a specific group of genes extends beyond the canonical termination site resulting in extended 3' UTR (*Ubr4*) or a possible fusion transcript by the transcription encroachment of a downstream gene (*Abca5*) (Figure 6.2) (Mahat et al., 2016). Because cancer cells are under proteotoxic stress (Dai et al., 2012), transcription read-through in stressed cells and cancers likely arises from a common mechanism that goes awry in both conditions. Examining the mechanism and machinery involved in transcription read-through during the stress response offers critical insights on the aberrant molecular mechanism and provides new set of targets for cancer therapeutics.

Characterize the RNA species generated by transcription read-through

RNAs with extended 3'-region and fusion transcripts generated by transcription read-through are characterized by assessing nascent transcription in untransformed human cell lines prior to stress, during stress, and after recovery from stress using enhanced PRO-seq (ePRO-seq,

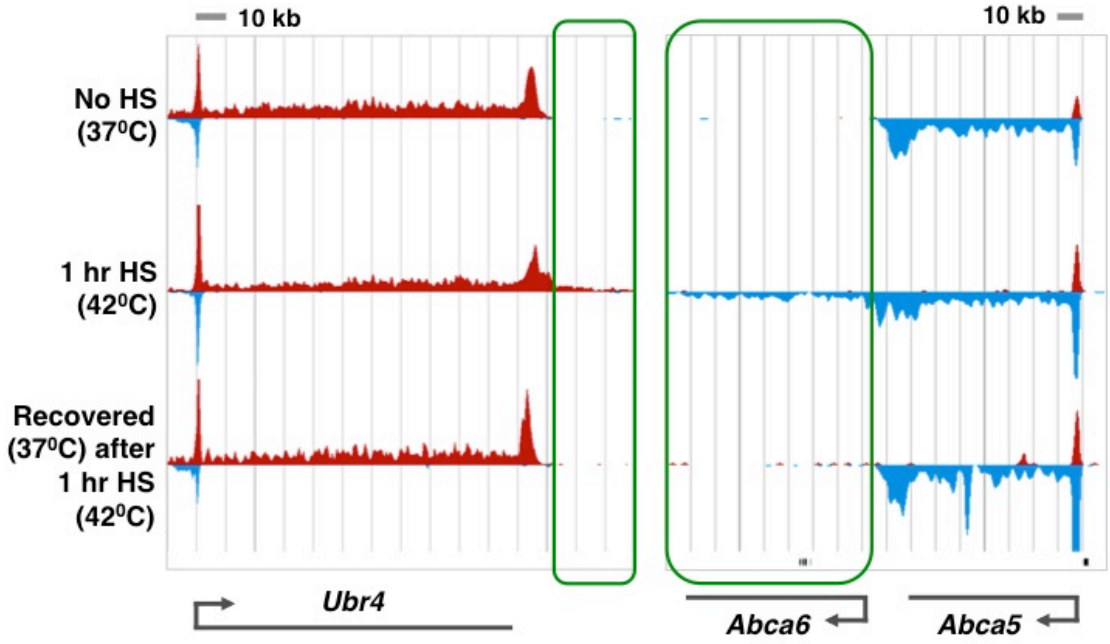


Figure 6.2. Transcription read-through in heat stress
Precision nuclear run-on sequencing (PRO-seq) in MEFs showing transcription read-through beyond transcription termination site. The arrows indicate direction of genes and PRO-seq density in plus and minus strands are shown in red and blue respectively. Green boxes highlight transcription read-through.

manuscript in prep). The translational output of these RNAs are evaluated using ribosome profiling (Ingolia et al., 2009), and the temporal control and efficiency of translation of transcription read-through products are elucidated.

High-throughput CRISPR/Cas9 & plasmid cDNA library screens to identify factors regulating transcription read-through

To identify factors responsible for transcription read-through, a reporter gene for qualitative evaluation of transcription read-through is created by inserting a GFP cDNA in-frame in the intergenic region downstream of a gene that undergoes transcription read-through upon stress (identified in Aim 1). After stably expressing the reporter gene and Cas9 in a cell line, cells are transfected with genome-scale lentiviral single-guide RNA (sgRNA) library (Wang et al., 2014). Lack of GFP expression after heat stress (FACS-sorted non-fluorescent cells) indicates the suppression of transcription read-through. Targeted sgRNA sequencing in the non-fluorescent cells identifies factors essential for transcription read-through. Conversely, transfecting plasmid cDNA library (Wan et al., 2006) and targeted sequencing of transfected cDNA in non-fluorescent cells identifies factors whose limitation induce transcription read-through.

HSR in everyday life

Beyond the sphere of the molecular mechanism of HSR presented in this dissertation and the three aforementioned areas of stress response with

direct relevance to human health, other relatively unexplored sides of stress response are equally intriguing. Acute and chronic exercise increases the expression of HSP72 (Leon, 2016; Yamada et al., 2008) and exercise-induced transient increase in HSP levels restricts vascular inflammation (Noble and Shen, 2012). Nutrients and growth hormone mediated activation of stress response, which we inadvertently discovered (Chapter 5), provide novel insights on HSF1-independent induction of *HSPs*. Exposure of worms and flies to mild stress confers health benefits and increases their lifespan (Le Bourg et al., 2001; Rattan, 2005). These aspects of HSR are primed for in-depth exploration.

Perspective

HSR is an old play, running on broadway for decades. Everyone who visited the theater was enamored by the lead actor. She was fast, she was furious, and she had a dramatic way of bringing changes that mattered to the life of the play. There was no debate that she was the heart and soul of the show. But for the longest time, everyone believed that this was one-man show. Then times changed. Slowly and steadily, people got closer and closer to the stage. Rumors started that the play was more complex than the role of the lead actor. Finally, one day, someone who had the front-row seat climbed to the stage and lifted the curtain. And all of sudden, right then and there, the cast and the crew behind the lead actor came to light. The show was definitely

complex, a lot more complex than anyone had imagined. People soon found out that the lead actor was not as powerful as she was portrayed to be.

Since then, a lot of new information has come out about the show. It apparently is a very intriguing show. Many folks walk out of the theater believing that they now know everything about the show. They seem content. They are happy. But the one with a smile is the old lady by the door of the theater, fixing her worn-out mat to sleep under the sky. She knows that there is still a lot more in the show that the folks don't know. She knows the play has more players and multiple layers. She also knows that someday someone will get even closer than anyone ever has been and reveal more secrets. But for today, she quietly slips under her blanket and whispers – "the show goes on!"

REFERENCES

Choi, E.-S., Lee, H., Lee, C.-H., and Goh, S.-H. (2016). Overexpression of KLHL23 protein from read-through transcription of PHOSPHO2-KLHL23 in gastric cancer increases cell proliferation. FEBS Open Bio *6*, 1155–1164.

Core, L.J., Waterfall, J.J., and Lis, J.T. (2008). Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. Science *322*, 1845–1848.

Dai, C., Dai, S., and Cao, J. (2012). Proteotoxic stress of cancer: Implication of the heat-shock response in oncogenesis. Journal of Cellular Physiology *227*, 2982–2987.

Danko, C.G., Hah, N., Luo, X., Martins, A.L., Core, L., Lis, J.T., Siepel, A., and Kraus, W.L. (2013). Signaling pathways differentially affect RNA polymerase II initiation, pausing, and elongation rate in cells. Molecular Cell *50*, 212–222.

Grosso, A.R., Leite, A.P., Carvalho, S., Matos, M.R., Martins, F.B., Vítor, A.C., Desterro, J.M.P., Carmo-Fonseca, M., and de Almeida, S.F. (2015). Pervasive transcription read-through promotes aberrant expression of oncogenes and RNA chimeras in renal carcinoma. Elife *4*, 43.

Ingolia, N.T., Ghaemmaghami, S., Newman, J.R.S., and Weissman, J.S. (2009). Genome-Wide Analysis in Vivo of Translation with Nucleotide Resolution Using Ribosome Profiling. Science *324*, 218–223.

Kim, J., Kim, S., Ko, S., In, Y.-H., Moon, H.-G., Ahn, S.K., Kim, M.K., Lee, M., Hwang, J.-H., Ju, Y.S., et al. (2015). Recurrent fusion transcripts detected by whole-transcriptome sequencing of 120 primary breast cancer samples. Genes Chromosomes Cancer *54*, 681–691.

Kwak, H., Fuda, N.J., Core, L.J., and Lis, J.T. (2013). Precise maps of RNA polymerase reveal how promoters direct initiation and pausing. Science *339*, 950–953.

Le Bourg, É., Valenti, P., Lucchetta, P., and Payre, F. (2001). Effects of mild heat shocks at young age on aging and longevity in Drosophila melanogaster. Biogerontology *2*, 155–164.

Leon, L.R. (2016). Common mechanisms for the adaptive responses to exercise and heat stress. J. Appl. Physiol. *120*, 662–663.

Lis, J.T., Mason, P., Peng, J., Price, D.H., and Werner, J. (2000). P-TEFb kinase recruitment and function at heat shock loci. Genes & Development *14*, 792–803.

Mahat, D.B., Salamanca, H.H., Duarte, F.M., Danko, C.G., and Lis, J.T. (2016). Mammalian Heat Shock Response and Mechanisms Underlying Its Genome-wide Transcriptional Regulation. Molecular Cell *62*, 63–78.

Nacu, S., Yuan, W., Kan, Z., Bhatt, D., Rivers, C.S., Stinson, J., Peters, B.A., Modrusan, Z., Jung, K., Seshagiri, S., et al. (2011). Deep RNA sequencing analysis of readthrough gene fusions in human prostate adenocarcinoma and reference samples. BMC Med Genomics *4*, 11.

Noble, E.G., and Shen, G.X. (2012). Impact of Exercise and Metabolic Disorders on Heat Shock Proteins and Vascular Inflammation. Autoimmune Diseases *2012*, 1–13.

Rabindran, S.K., Haroun, R.I., Clos, J., Wisniewski, J., and Wu, C. (1993). Regulation of heat shock factor trimer formation: role of a conserved leucine zipper. Science *259*, 230–234.

Rattan, S. (2005). Hormetic modulation of aging and longevity by mild heat stress (Nonlinearity in Biology).

Santagata, S., Hu, R., Lin, N.U., Mendillo, M.L., Collins, L.C., Hankinson, S.E., Schnitt, S.J., Whitesell, L., Tamimi, R.M., Lindquist, S., et al. (2011). High levels of nuclear heat-shock factor 1 (HSF1) are associated with poor prognosis in breast cancer. Proceedings of the National Academy of Sciences

108, 18378–18383.

Valentijn, L.J., Koster, J., and Versteeg, R. (2006). Read-through transcript from NM23-H1 into the neighboring NM23-H2 gene encodes a novel protein, NM23-LV. Genomics *87*, 483–489.

Wan, K.H., Yu, C., George, R.A., Carlson, J.W., Hoskins, R.A., Svirskas, R., Stapleton, M., and Celniker, S.E. (2006). High-throughput plasmid cDNA library screening. Nat Protoc *1*, 624–632.

Wang, T., Wei, J.J., Sabatini, D.M., and Lander, E.S. (2014). Genetic Screens in Human Cells Using the CRISPR-Cas9 System. Science *343*, 80–84.

Yamada, P., Amorim, F., Moseley, P., and Schneider, S. (2008). Heat shock protein 72 response to exercise in humans. Sports Med *38*, 715–733.

Yoon, Y.J., Kim, J.A., Shin, K.D., Shin, D.-S., Han, Y.M., Lee, Y.J., Lee, J.S., Kwon, B.-M., and Han, D.C. (2011). KRIBB11 inhibits HSP70 synthesis through inhibition of heat shock factor 1 function by impairing the recruitment of positive transcription elongation factor b to the hsp70 promoter. Journal of Biological Chemistry *286*, 1737–1747.

Zuo, J., Baler, R., Dahl, G., and Voellmy, R. (1994). Activation of the DNA-binding ability of human heat shock transcription factor 1 may involve the transition from an intramolecular to an intermolecular triple-stranded coiled-coil structure. Mol. Cell. Biol. *14*, 7557–7568.

APPENDIX A[d]

BASE-PAIR RESOLUTION GENOME-WIDE MAPPING OF ACTIVE RNA POLYMERASES USING PRECISION NUCLEAR RUN-ON (PRO-seq)

Summary

We provide a protocol for precision nuclear run-on sequencing (PRO-seq) and its variant, PRO-cap, which map the location of active RNA polymerases (PRO-seq) or transcription start sites (TSSs) (PRO-cap) genome-wide at high resolution. The density of RNA polymerases at a particular genomic locus directly reflects the level of nascent transcription at that region. Nuclei are isolated from cells and, under nuclear run-on conditions, transcriptionally engaged RNA polymerases incorporate one or, at most, a few biotin-labeled nucleotide triphosphates (biotin-NTPs) into the 3' end of nascent RNA. The biotin-labeled nascent RNA is used to prepare sequencing libraries, which are sequenced from the 3' end to provide high-resolution positional information for the RNA polymerases. PRO-seq provides much higher sensitivity than ChIP-seq, and it generates a much larger fraction of usable

[d] Adapted from Mahat, D. B.*, Kwak, H.*, Booth, G. T., Jonkers, I. H., Danko, C. G., Patel, R. K., Waters, C.T., Munson, K., Core, L.J., & Lis, J.T. (2016). Base-pair-resolution genome-wide mapping of active RNA polymerases using precision nuclear run-on (PRO-seq). *Nature Protocols*, *11*(8), 1455–1476. *denotes equal contribution. Contribution is shown in each figure. Reprinted with permission from Nature Publishing.

sequence reads than ChIP-seq or NET-seq (native elongating transcript sequencing). Similarly to NET-seq, PRO-seq maps the RNA polymerase at up to base-pair resolution with strand specificity, but unlike NET-seq it does not require immunoprecipitation. With the protocol provided here, PRO-seq (or PRO-cap) libraries for high-throughput sequencing can be generated in 4–5 working days. The method has been applied to human, mouse, *Drosophila melanogaster* and *Caenorhabditis elegans* cells and, with slight modifications, to yeast.

Introduction

The ability to measure the density of RNA polymerase across the genome provides a comprehensive and quantitative snapshot of transcription(Fuda et al., 2009). Collecting a series of these snapshots in response to regulatory switches reveals the identity of genes that respond immediately or secondarily to specific signals, and provides critical insights to the mechanisms of their regulation(Adelman and Lis, 2012). Quantifying RNA polymerase density along the genes is also critical for deciphering the regulatory steps involved in transcription.

In addition to protein coding genes, many other regions in the genome (such as upstream divergent regions, regions downstream of mRNA poly A sites, and enhancers) are transcribed to various extents. Enhancers produce short unstable RNAs (eRNAs) that do not encode proteins(Core et al., 2014)

but delineate major hubs of transcription regulation(Heinz et al., 2015). Differential regulation of enhancer mediated transcription is implicated in various diseases(Vahedi et al., 2015), and understanding this regulation is important for deciphering the transcriptional changes in response to developmental, nutritional, and environmental cues. However, sequencing of total RNA by RNA-seq is inefficient in detecting these unstable RNAs.

A number of methods have been described that enrich and sequence nascent RNAs associated with RNA polymerase. These methods are either based on immunoprecipitation of RNA polymerase(Churchman and Weissman, 2011; Larson et al., 2014; Nojima et al., 2015) or are dependent on purification of insoluble chromatin(Weber et al., 2014). Therefore, these methods are highly dependent on antibody specificity or the purity of the chromatin fraction respectively. We have developed nuclear run-on based methods to map active RNA polymerases and their start sites genome-wide at up to base pair resolution(Core et al., 2008; Kwak et al., 2013). In these methods, the endogenous activity of RNA polymerase is used to selectively label nascent RNAs. The ability to affinity-purify nuclear run-on RNA multiple times during the course of library preparation provides an approximate million-fold enrichment of the nascent RNA over other forms of RNA and thereby effectively eliminates background(Core et al., 2008). Furthermore, because the RNA is sequenced, the direction of transcription can be unambiguously identified.

Development of PRO-seq

PRO-seq is based on Global Run-On sequencing (GRO-seq), a genome-wide adaptation of nuclear run-on assays that have been used classically to measure transcription of target genes. In GRO-seq, bromouridine-labeled nascent RNAs are affinity-purified and analyzed by high-throughput sequencing to map RNA polymerase positions. Extremely high sensitivity and specificity is achieved through multiple distinct affinity-purification steps(Core et al., 2008; Jonkers et al., 2014). GRO-seq uses bromouridine as the substrate for the nuclear run-on reaction, enabling RNA polymerase to add multiple nucleotides to the nascent RNA. Therefore, the resolution of GRO-seq is tens of bases.

However, to understand the molecular mechanisms of transcriptional elongation and promoter-proximal pausing, RNA polymerase mapping at base-pair resolution is required. Such resolution enables mechanistic modeling of how DNA sequences, nucleosomes, or other DNA-binding factors affect RNA polymerase elongation and gene expression(Adelman and Lis, 2012; Kwak and Lis, 2013). To achieve base-pair resolution, we used a modified nuclear run-on that limits the number of labeled nucleotides added to the nascent RNA(Core et al., 2008; Hah et al., 2011; Kwak et al., 2013; Larschan et al., 2011; Min et al., 2011). In PRO-seq, biotin-labeled nucleotide triphosphates are provided as the substrates for the nuclear run-on reaction. The incorporation of a biotin-NTP by an RNA polymerase inhibits further

incorporation of biotin-NTPs into the nascent RNA. Sequencing from the 3' end of the nascent transcript therefore identifies the last incorporated NTP and reveals the precise location of the active site of the RNA polymerase engaged with its nascent RNA.

Identification of the precise position of transcription start sites (TSSs) is also important to understand how DNA elements, general transcription factors and transcription activators recruit RNA polymerase to genes and enhancers. RNA polymerase initiates transcription at TSS and quickly transcribes a short region before pausing at a promoter-proximal site. However, because the nascent transcripts are sequenced from the 3' end in PRO-seq, the positional information of where RNA polymerase began transcription is mostly lost. We therefore developed PRO-cap by modifying the sequencing strategy of PRO-seq to sequence the capped nascent RNA from 5' end, enabling transcription start sites to be identified at the RNA synthesis level(Kwak et al., 2013).

Overview of the procedure

A general overview of the PRO-seq and PRO-cap experimental procedures is shown in Figure A.1. Nuclei from cells are rapidly isolated and native nucleotides are washed away to halt transcription. However, RNA polymerases remain engaged on the DNA and retain their enzymatic activity. Incubation of isolated nuclei with biotin-labeled nucleotide triphosphates allows the RNA polymerases to actively elongate and label the nascent RNA. For

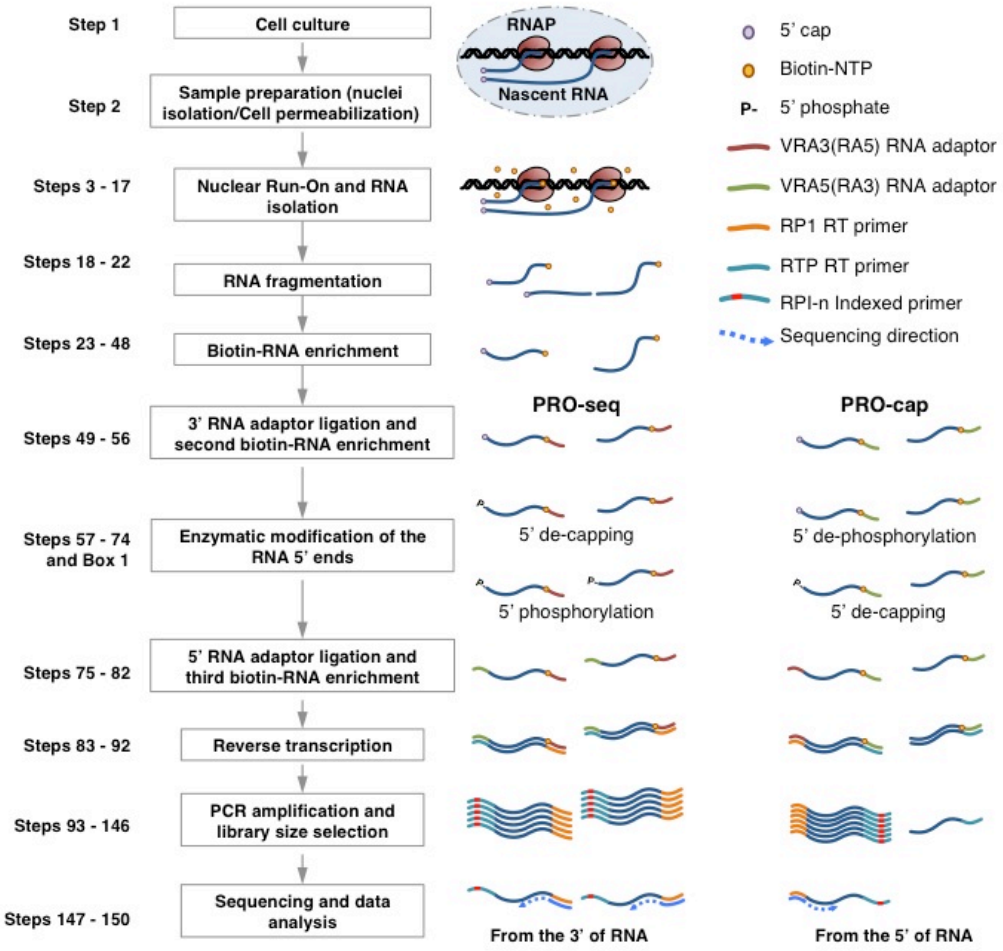


Figure A.1. Flowchart of the PRO-seq and PRO-cap protocol
PRO-seq and PRO-cap share common steps, but are different at the adaptor ligation, reverse transcription, and 5' modification stages. In PRO-seq, the 3' RNA adaptor is modified to have the reverse complement sequence of a standard 5' RNA adaptor. Similarly, the 5' RNA adaptor and the reverse transcription primer are modified for PRO-seq. This allows 3' sequencing of the RNA on a standard Illumina platform. For PRO-cap, standard small RNA adaptors are used, and result in a 5' sequencing as usual. In the 5' modification step, PRO-seq RNAs are de-capped and re-phosphorylated to ensure all RNA is accessible for 5' RNA adaptor ligation. PRO-cap RNAs are de-phosphorylated first to convert all forms of 5' end except for the 5' capped ends to a 5' hydroxyl end, in order to restrict the ligation of 5' adaptor only to the cap containing nascent RNAs after cap removal. Then the 5' caps are converted to 5' monophosphates using the tobacco acid pyrophosphatase, and can be ligated to a 5' RNA adaptor.

PRO-seq only, the labeled nascent RNA is hydrolyzed with NaOH to generate RNA fragments suitable for sequencing (~100 bp in length). The RNA containing a biotin nucleotide is then enriched by affinity purification using streptavidin-coated magnetic beads. The biotin-streptavidin interaction is very stable ($K_d \sim 10^{-14}$ mol/L), which allows stringent washing of the magnetic beads to minimize contamination with unlabeled RNA. A 3' sequencing adapter is then ligated to the hydroxyl (OH) group at the 3'end of nascent RNA followed by another affinity purification to further enrich nascent RNA and remove unligated adapter sequences. Preparation for 5' sequencing adaptor ligation differs for PRO-Seq and PRO-cap. For PRO-seq, the 5' cap is removed from unhydrolyzed short nascent RNA using either Tobacco acid pyrophosphatase (TAP) or RNA 5' Pyrophosphohydrolase (RppH). The 5' OH generated by base hydrolysis is then converted to 5' phosphate by treatment with T4 polynucleotide kinase (PNK) For PRO-cap, uncapped RNA with a 5'-monophosphate is degraded using 5´-Phosphate-dependent exonuclease. 5' triphosphates and monophosphates are removed from any remaining contaminating uncapped RNA with Alkaline phosphatase. Only then is the 5' cap of nascent RNA removed with TAP or RppH treatment. After these chemical modifications, a 5' sequencing adapter is ligated to the nascent RNA and a third round of affinity purification performed to enrich for nascent RNA with sequencing adapters on both ends. Nascent RNA is then reverse transcribed and test-PCR-amplified to determine an appropriate number of

PCR cycles for final amplification; this latter step is critical to avoid over amplification. During the final amplification, barcodes can be added so that multiple libraries can be pooled for sequencing. Finally, the PCR amplified libraries are size-selected for a range of 140-350 bp and sent for high-throughput sequencing. Sequencing depth of 25-50 million for mammalian cells, 10-20 million for organisms with smaller genome size such as Drosophila, and 5-10 million for yeast cells provide useful information. Relatively short read length such as 40-50 bp is sufficient. The sequencing result is a text-based list of short nucleotide sequences and its sequencing quality parameters are provided in a 'fastq' format(Cock et al., 2010). The sequences may contain varying lengths of adapter sequences, which are removed, and the adaptor-removed sequences are then aligned to the appropriate genome. Finally, the aligned sequences are used to generate coverage files that can be used to visualize and analyze the data.

Advantages and limitations of PRO-seq

The key advantages of PRO-seq are:

- It provides base-pair resolution and strand-specific information of global RNA polymerase occupancy.
- Background RNA contamination is hugely reduced due to almost a million fold purification of biotinylated nascent RNAs.

- It is highly sensitive in detecting rare and common nascent RNAs with a large dynamic range (>10^5).
- It can identify short unstable nascent RNAs transcribed from enhancer regions.

PRO-seq also has a number of limitations that should be considered when deciding which genome-wide RNA polymerase mapping strategy to use.

- In principle, PRO-seq results are ensemble profiles of potentially heterogeneous populations of cells - and this is generally true for all multi-cell, high-throughput sequencing analyses. Unlike mature RNA molecules that are present in multiple copies per cell, RNA polymerase at a specific genomic position can only yield at most two copies of nascent RNA. So while it may be possible to adapt PRO-seq to measure nascent transcript levels for abundantly expressed genes in single cells, genome-wide mapping of RNA polymerase in single cells using PRO-seq would remain a challenge.
- PRO-seq detects only the active RNA polymerase and so RNA polymerases in the pre-initiation complex will not be detected. There is also a possibility that other forms of stalled RNA polymerases, such as back-tracked polymerases, may not be detected, although nuclear run-on conditions may allow some of these polymerases to realign the active site through thermal motion. Generally, the signals seen by ChIP-seq of RNA Polymerase II (Pol II) and our genome-wide run-on methods quantitatively

agree(Core et al., 2012), so the bulk of Pol II is detectable by GRO- and PRO-seq methods.

- Compared to GRO-seq that adds a longer extension to the 3' end of nascent RNA(Core et al., 2008), PRO-seq only adds one or a few nucleotides in order to provide higher resolution mapping. However, there is a possibility that RNA polymerases positioned very close to the TSS may not be detected because the nascent RNA may not be long enough to be uniquely mapped to the genome. In this case, GRO-seq may provide more accurate quantification of promoter proximal RNA polymerases. Likewise, RNA polymerases positioned in a repetitive sequence region of the genome is difficult to unambiguously map to a particular repeat.
- PRO-seq does not distinguish nascent transcription derived from different RNA polymerases (Pol I vs Pol II vs Pol III) unless carried out in the presence of inhibitors of specific RNA polymerases. Also, unlike NET-Seq, nascent RNA associated with specific RNA polymerase modifications (e.g. phosphorylations of the C-terminal domain) cannot be selectively detected using PRO-seq(Nojima et al., 2015).

Applications of PRO-seq and PRO-cap

The most common use of PRO-seq is for the analysis of genome-wide transcription levels with directional information and with higher resolution and sensitivity than an RNA polymerase ChIP-seq assay. PRO-seq provides an

independent layer of gene expression analysis distinct from mRNA-seq, revealing the transcriptional stages of regulation before the influence of mRNA processing or stability control. The increased resolution and the directional information become useful in distinguishing upstream divergent (also called upstream antisense) transcription(Core et al., 2008; Seila et al., 2008).

PRO-cap(Kwak et al., 2013) can capture TSS at the nascent RNA synthesis level in contrast to other TSS analyses that use mature RNA(Andersson et al., 2014; Carninci et al., 1996; Forrest et al., 2014). This becomes an advantage in detecting enhancer transcripts(Core et al., 2014; Hah et al., 2013; Wang et al., 2011), upstream antisense transcription(Core et al., 2008; 2014), or other types of unstable transcripts, and avoiding post-transcriptional background capping events(Fejes-Toth et al., 2009).

Alternatives to PRO-seq

RNA polymerase can be mapped genome-wide by a variety of different strategies.

- *ChIP-seq*: In this approach, RNA polymerase proteins are cross-linked to the DNA and then RNA polymerase II (Pol II) is purified by immuno-precipitation. Pol II associated DNA is identified and quantified by high-throughput sequencing thereby providing an estimate of the amount of Pol II at different sites on the genome. The resolution of ChIP is usually limited by the size of the fragmented DNA in the chromatin at the

immunoprecipitation step. A variant of this method called ChIP-exo overcomes the resolution limitation by additionally treating the DNA fragments from the Pol II ChIP with a DNA exonuclease(Rhee and Pugh, 2012). The exonuclease digests DNA from the 3' end of both strands, stopping near the cross-linked polymerase complex. An additional limitation of ChIP-seq is that Pol II-unbound genomic regions that interact with a Pol II-bound region through three-dimensional looping can be falsely identified with this method due to the use of cross-linking. Finally, ChIP-seq will map all forms of Pol II, including transcriptionally inactive Pol IIs and Pol IIs in antisense orientation: therefore, the direction of the transcription is not directly disclosed.

- *Permanganate footprinting*: this method can be used to identify the single stranded DNA created by the transcription bubble formed on DNA by the RNA polymerase. The non-template strand of the DNA is exposed to single-strand specific breakage at T residues through a series of chemical treatments. A method called Permanganate-ChIP-seq couples permanganate footprinting to Pol II ChIP and thereby maps the cleaved ends of the DNA from the single-stranded region of transcription bubbles(Li et al., 2013). This directly maps the transcription active site with high resolution. Permanganate mapping depends on the presence of thymine residues in the non-templated single-stranded DNA of the bubble that are not masked by protein binding. Although the protocol enriches for Pol II in a

single chromatin immunoprecipitation, other regions that expose single stranded thymine such as in other DNA-RNA hybrids or intra-strand DNA hairpins could contribute to background.

- *Native Elongating Transcript sequencing (NET-seq):* a number of chromatin bound nascent RNA based methods, including NET-seq and its variants, have been developed for mapping RNA polymerase(Churchman and Weissman, 2011; Mayer et al., 2015; Nojima et al., 2015; Weber et al., 2014). In the original NET-seq protocol(Churchman and Weissman, 2011), the RNA polymerase complex is immunoprecipitated and the co-purified native RNA is sequenced. The 3' end of the nascent RNA provides base-pair resolution mapping of RNA polymerase. This method is ideally suited to examine the occupancy of differently modified RNA polymerases. In practice, the efficiency of NET-seq relies on the degree of enrichment provided by the single immuno-precipitation step. Because the method detects the 3' ends of all RNAs that are associated with Pol II, it also captures 3' ends of intermediates of co-transcriptional splicing and micro-RNA production that can complicate mapping of transcriptionally-engaged Pol II(Nojima et al., 2015).

Experimental design

Cells. In our lab, we have successfully generated PRO-Seq libraries for cells from plant (unpublished, G.T.B.), yeast (unpublished, G.T.B.), *Drosophila*(Kwak et al., 2013), and mammals(Core et al., 2014; Mahat et al., 2016). In general, the higher the number of cells the better the PRO-seq read coverage of the genome. However, a minimum of 5-10 million nuclei or permeabilized cells is required for a single PRO-seq library regardless of cell type. In principle, the application of PRO-seq in yeast, including *S. pombe* and *S. cerevisiae*, is very similar to that of other organisms; however, some alterations are required in yeast cell permeabilization(García-Martínez et al., 2004), run-on reaction, and post run-on RNA extraction procedures(Collart and Oliviero, 2001); required modifications for yeast are indicated in the appropriate steps of the Procedure.

Sample preparation. Isolation of nuclei for nuclear run-on is a critical step in the procedure not only to preserve the enzymatic activity of the RNA polymerase, but also to capture the precise position of the RNA polymerase on genes. Starting with 10-20 million cells per library is recommended considering the efficiency of nuclei isolation process (~50%). The whole process should take place in the cold room on ice as far as is possible. Isolated nuclei can be resuspended in the glycerol-containing storage buffer, and quickly frozen in liquid nitrogen for long term storage at -80°C. We have used permeabilized cells in PRO-seq as an alternative to isolating nuclei, making handling easier

and reducing loss of sample; cell permeabilization has a much higher efficiency (~90%) than nuclei isolation. Cell permeabilization conditions may differ between cell types and may need to be optimized; we provide a general method for permeabilization in the Procedure as well as a version optimized for yeast cells.

Spike-in for library normalization. Disproportionate loss of RNA and/or cDNA can occur during multiple stages of the PRO-seq library preparation, which spans 4-5 days and involves several handling steps. Even with the use of identical starting material, uneven loss of libraries could affect the genome-wide RNA polymerase density between libraries. To control for handling effects on library yield, a small fraction (1-5%) of cells with a distinct genome can be added during library preparation; adding an identical number of spike-in cells to different libraries enables normalization between different conditions. We have used *Schizosaccharomyces pombe* to normalize *Saccharomyces cerevisiae* and vice versa, and *Drosophila* cells to normalize mammalian cells and vice versa. When using the cell-permeabilization approach, spike-in cells should be added and permeabilized together with the experimental cells. For the nuclei isolation approach, spike-in cells should be added to the experimental cells prior to nuclei isolation and dounced together.

Nuclear run-on. In PRO-seq, biotin-NTPs are used as the nuclear run-on substrates. The K_m of each of the NTPs as substrates for RNA polymerase lie in the range of 1–20 μM(Job et al., 1984). Therefore, final substrate

concentration greater than 1-20 μM range (~25 μM) is, in general, sufficient for each biotin-NTP substrate. Depending on the purpose of the experiment, biotin-NTP substrates can be supplied in different combinations: individual biotin run-on, 4 biotin run-on, 2 biotin run-on or 1 biotin run-on.

- *Individual-biotin run-on*: To obtain the most precise mapping of the RNA polymerase, four separate PRO-seq libraries are made, each supplied with only one type of biotin-NTP in the run-on reaction. This ensures that the RNA polymerase adds only one, or at most a few (when the polymerase is positioned at multiple stretches of same nucleotide) biotin-NTPs to the nascent RNA. In this case, 4 times more sample is required.
- *4-Biotin run-on*: we found that all 4 biotin-NTPs can be supplied in a single reaction and the Pol II only incorporates one or at most a few bases, giving an equivalent resolution to single biotin run-on. The reason for this is unclear, but we speculate that steric hindrance in the active site of RNA polymerase prevents incorporation of multiple biotinylated nucleotides.
- *2-Biotin run-on*: When the amount of sample or the cost is limiting, unlabeled NTPs can be used in combination with biotin NTPs. Use of biotinylated purine nucleotides (biotin-ATP, biotin-GTP) is more costly than that of the pyrimidine nucleotides (biotin-CTP, biotin-UTP). A combination of biotin-CTP, biotin-UTP, ATP, and GTP can be supplied to the nuclear run-on reaction, providing reasonable resolution and cost.

- *1-Biotin run-on*: If a longer run-on RNA is preferred, combinations of biotin-CTP with unlabeled CTP, UTP, ATP, and GTP can be used effectively in a biotin-NTP form of GRO-seq. This approach can be useful for increasing sequencing coverage of RNA polymerases that reside near the TSSs. While most transcriptionally-engaged RNA polymerases near the 5' ends reside between 30-60 nucleotides from the TSS(Kwak et al., 2013), RNA polymerases closer to the TSS may fail to map uniquely. Additionally, the longer run-on extensions of nascent RNAs may be desired for distinguishing allele-specific nascent transcription.

PCR amplification of PRO-seq library. When the number of cells and/or nuclei are limited for PRO-seq library preparation, a higher number of PCR amplification cycles will be required to generate sufficient library for reliable quantification and accurate loading into the sequencer. However, a higher number of PCR cycles can result in amplification bias of some sequences. To avoid PCR-induced biases, molecular barcodes(Fu et al., 2014) can be introduced as part of the 3' RNA adapter, which is ligated to the nascent RNAs. Duplicate reads generated by PCR over-amplification can be identified by identical barcodes and computationally filtered at the stage of mapping the sequenced reads to the genome.

Materials

Reagents

CRITICAL: Extreme care should be taken to avoid nuclease contamination. Use nuclease-free reagents and change gloves routinely.

- Appropriate cell line(s) e.g. K562, GM12878, MCF7, Hela, embryonic stem cells, MEFs, mouse 3T3 cells, *Drosophila* S2, yeast.

  CAUTION: Before use, cells should be checked for contamination.

Chemical stocks

- Diethyl pyrocarbonate (DEPC; Sigma-Aldrich, cat. no. D5758)

  CAUTION: DEPC is toxic and harmful. Proper eyeshield, faceshield, full-face respirator, and gloves are required while handling DEPC.
- Sodium chloride, NaCl (Sigma-Aldrich, cat. no. S9888)
- Potassium chloride, KCl (Avantor, cat. no. 6858-04)
- Magnesium chloride, $MgCl_2$ (Avantor, cat. no. 5958-04)
- Sucrose (Sigma-Aldrich, cat. no. S0389)
- Calcium chloride, $CaCl_2$ (Sigma-Aldrich, cat. no. C1016)
- Magnesium acetate, $MgAc_2$ (Sigma-Aldrich, cat. no. M5661)
- Ammonium acetate, NH_4Ac (Sigma-Aldrich, cat. no. A1542)
- Sodium acetate NaOAc (Sigma-Aldrich, cat. no. S2889)
- EDTA (Sigma-Aldrich, cat. no. E9884)
- EGTA (Sigma-Aldrich, cat. no. E3889)
- Protease inhibitor cocktail, EDTA-free (Roche, cat. no. 11873580001)

- Sodium hydroxide, NaOH (Avantor, cat. no. 7708-10)
- Triton X-100, (Calbiochem, cat. no. 9410)
- Nonidet P40 (NP40) Substitute, (Sigma-Adlrich, cat. no. 11332473001)
- Sarkosyl (Sigma-Aldrich, cat. no. L5125)
- Tween-20 (Sigma-Aldrich, cat. no. P9416)
- Phosphate buffer saline, PBS pH 7.4 (Gibco, cat. no. 10010031).
- TRIS (Avantor, cat. no. 4109-02)
- Hydrochloric acid, HCl (Avantor, cat. no. 4613-05)
- DTT (Sigma-Aldrich, cat. no. D0632)
- Betaine (Sigma-Aldrich, cat. no. B0300)
- Glycerol (Sigma-Aldrich, cat. no. G5516)

Biotin Nuclear Run-On and enrichment

- Biotin-11-ATP (PerkinElmer, cat. no. NEL544001EA)
- Biotin-11-CTP (PerkinElmer, cat. no. NEL542001EA)
- Biotin-11-GTP (PerkinElmer, cat. no. NEL545001EA)
- Biotin-11-UTP (PerkinElmer, cat. no. NEL543001EA)
- ATP, 10mM (Roche, cat. no. 11 277 057 001)
- GTP, 10mM (Roche, cat. no. 11 277 057 001)
- UTP, 10mM (Roche, cat. no. 11 277 057 001)
- P-30 column, RNase free (BIORAD, cat. no. 732-6250)
- Streptavidin M280 beads (Invitrogen, cat. no. 112.06D)

Reagents for nucleic acid extraction

- Trizol (Ambion, cat. no. 115596018)

  CAUTION: Trizol is harmful and contact with skin, eye or inhalation should be avoided. Use it inside a fume hood.
- Trizol LS (Ambion, cat. no. 10296028)

  CAUTION: Trizol is harmful and contact with skin, eye or inhalation should be avoided. Use it inside a fume hood.
- Chloroform (Calbiochem, cat. no. 3150)
- GlycoBlue (Ambion, cat. no. AM9515)
- Ethanol, 100% (PHARMCO-AAPER, cat. no. 111000200)
- Ethanol, 75%(vol/vol)
- Phenol:Chloroform, Tris buffered (Thermo Scientific, cat. no. 17909)

  CAUTION: Phenol:Chloroform is harmful and contact with skin, eye or inhalation should be avoided. Use it inside a fume hood.
- Phenol, (Ambion, cat. no. 9700).

  CAUTION: Phenol is harmful and contact with skin, eye or inhalation should be avoided. Use it inside a fume hood.

Enzymes and recombinant protein reagents

- RNase inhibitor, 40 units/μl (Ambion, cat. no. AM2696)
- T4 RNA ligase I, 10 units/μl (NEB, cat. no. M0204). Supplied with 10× T4 RNA ligase buffer, 10 mM ATP, and PEG, 50%(wt/vol).
- 5´-phosphate-dependent exonuclease, 1 unit/μl (Epicenter, cat. no. TER51020) (required for PRO-cap only). Supplied with 10× rxn buffer A.

- Alkaline phosphatase, 10 units/μl (NEB, cat. no. M0290) (required for PRO-cap only). Supplied with 10× Alkaline phosphatase buffer. Alternatively, Antarctic phosphatase, 5 units/μl (NEB, cat. no. M0289) can be used.
- Tobacco acid pyrophosphatase, 10 units/μl (TAP) (Epicenter, cat. no. T19500). Supplied with 10× TAP buffer. Alternatively, RNA 5' Pyrophosphohydrolase, 5 units/μl (RppH) (NEB, cat. no. M0356S) can be used with ThermoPol Reaction buffer (NEB, cat. no. B9004S).
- T4 polynucleotide kinase, 10 units/μl (PNK) (NEB, cat. no. M0201) (required for PRO-seq only). Supplied with10× PNK buffer.
- Superscript III reverse transcriptase (Invitrogen, cat. no. 56575). Supplied with 5× first strand buffer, and 0.1M DTT.
- dNTP mix, 12.5 mM each (Roche, cat. no. 03 622 614 001)
- Phusion polymerase, 2 units/μl (NEB, cat. no. M0530). Supplied with 5× High-Fidelity buffer.
- RNA and DNA oligos. (Custom synthesis from IDT DNA, RNase-free HPLC purified) See Table 1 and Reagent Setup for details. Further information about barcoding and sequencing indexes can be found at http://support.illumina.com/content/dam/illumina-support/documents/documentation/chemistry_documentation/samplepreps_truseq/truseqsampleprep/truseq-library-prep-pooling-guide-15042173-01.pdf

Electrophoresis

- DNA grade agarose (BioRad, cat. no. 161-3102EDU)
- Tris/Acetic acid/EDTA (TAE), 50× (BioRad, cat. no. 161-0773). Alternatively, a 50x TAE buffer can be made in house (see Reagent Setup section)
- Glacial acetic acid (Sigma-Aldrich, cat. no. 537020) for in house preparation of 50x TAE.
- Gel loading dye, Orange G 6× (NEB, cat. no. B7022S)
- 100 bp DNA ladder (Life Technologies, cat. no. 15628-019)
- 10 bp DNA ladder (Life Technologies, cat. no. 10821-015). Alternatively, 25 bp DNA ladders (ThermoFisher, cat. no. 10597011) can be used.
- SYBRGold nucleic acid gel stain, 6× (Life Technologies, cat. no. S-11494)
- Acrylamide (Protogel), 30%(wt/vol) (National Diagnostics, cat. no. EC-980). 30% acrylamide/methylene bisacrylamide solution from other sources is also compatible.
- TEMED (BioRad, cat. no. 161-0800)
- Ammonium Persulfate (APS), 10%(wt/vol) (BioRad, cat. no. 161-0700). Dissolve in H_2O.
- Tris/Boric acid/EDTA (TBE), 10× (BioRad, cat. no. 161-0770). Alternatively, a 10× TBE buffer can be made in house (see Reagent Setup section).
- Boric acid (Sigma-Aldrich, cat. no. B6768) for in house preparation of 10x TBE.

Equipment

- 2 heat blocks, one set at 37°C and the other at 65°C, each filled with water equilibrated at the appropriate temperature.
- Dounce homogenizer (Wheaton scientific, cat. no. 357546)
- Magnetic separator for Streptavidin magnetic beads (Invitrogen, cat. no. K1585-01)
- Rotating stand (Thermo Barnstead Labquake rotator, cat. no. 415110)
- Refrigerated centrifuge (Eppendorf, cat. no. 5417R)
- Microcentrifuge (Eppendorf, cat. no. 5415D)
- Speed vac dryer (Thermo Scientific, cat. no. 20-548-134)
- Dark Reader transilluminator (Clare Chemical Research, cat. no. DR89X)

Reagent setup

CRITICAL: All reagents, solutions, and buffers should be made with DEPC-treated water

DEPC-H_2O Add 0.1%(vol/vol) DEPC to H_2O. Mix overnight then autoclave and filter-sterilize the solution with a 0.22 μm filter. DEPC-H_2O can be prepared in advance and stored at room temperature (25^0C) for up to a year.

CAUTION: DEPC is toxic and harmful. Proper eyeshield, faceshield, full-face respirator, and gloves are required while handling DEPC.

5M NaCl Dissolve 14.61 g NaCl in 50 ml H_2O with 0.1%(vol/vol) DEPC. Mix overnight and then autoclave and filter.

CRITICAL: 5M NaCl can be prepared in advance and stored at room temperature (25^0C) for up to a year.

4M KCl Dissolve 3.73 g KCl in 50 ml H_2O with 0.1%(vol/vol) DEPC, mix overnight then autoclave and filter-sterilize.

CRITICAL: 4M KCl can be prepared in advance and stored at room temperature for up to a year.

1M $MgCl_2$ Dissolve 4.76 g $MgCl_2$ in 50 ml H_2O with 0.1%(vol/vol) DEPC, mix overnight then autoclave and filter-sterilize.

CRITICAL: 1M $MgCl_2$ can be prepared in advance and stored at room temperature for up to a year.

1M Sucrose Dissolve 171.15 g Sucrose in 500 ml H_2O with 0.1%(vol/vol) DEPC, mix overnight then autoclave and filter-sterilize.

CRITICAL: 1M Sucrose can be prepared in advance and stored at room temperature for up to a year.

1M $CaCl_2$ Dissolve 5.55 g $CaCl_2$ in 50 ml H_2O with 0.1%(vol/vol) DEPC, mix overnight then autoclave and filter-sterilize.

CRITICAL: 1M $CaCl_2$ can be prepared in advance and stored at room temperature for up to a year.

1M $MgAc_2$ Dissolve 7.12 g $MgAc_2$ in 50 ml H_2O with 0.1%(vol/vol) DEPC, mix overnight then autoclave and filter-sterilize.

CRITICAL: 1M $MgAc_2$ can be prepared in advance and stored at room temperature for up to a year.

1M NH_4Ac Dissolve 3.85 g NH_4Ac in 50 ml H_2O with 0.1%(vol/vol) DEPC, mix overnight then autoclave and filter-sterilize.

CRITICAL: 1M NH_4Ac can be prepared in advance and stored at room temperature for up to a year.

1M NaOAc, pH 5.3 Dissolve 4.1 g NaOAc in 50 ml H_2O and pH to 5.3, add 0.1%(vol/vol) DEPC, mix overnight then autoclave and filter-sterilize.

CRITICAL: 1M NaOAc can be prepared in advance and stored at room temperature for up to a year.

0.5M EDTA Dissolve 29.22 g EDTA in 100 ml DEPC treated H_2O, then autoclave and filter-sterilize.

CRITICAL: 0.5M EDTA can be prepared in advance and stored at room temperature for up to a year.

0.1M EGTA Dissolve 19.02 g EGTA in 50 ml DEPC treated H_2O, then autoclave and filter-sterilize.

CRITICAL: 0.1M EGTA can be prepared in advance and stored at room temperature for up to a year.

1N NaOH Dissolve 2 g NaOH in 50 ml DEPC treated H_2O, filter-sterilize.

CRITICAL: 1N NaOH can be prepared in advance in 50 ul aliquots, stored at -80°C for up to a year. Use freshly thawed aliquot each time.

CAUTION: NaOH is corrosive and contact with skin, eye or inhalation should be avoided.

10% Triton X-100 Dissolve 5 ml of Triton X-100 in 45 ml DEPC H_2O and filter-sterilize.

CRITICAL: 10% Triton X-100 can be prepared in advance and stored at room temperature for up to a year.

10% NP40 Dissolve 5 ml of NP40 in 45 ml DEPC H_2O and filter-sterilize.

CRITICAL: 10% NP40 can be prepared in advance and stored at room temperature for up to a year.

2% Sarkosyl Dissolve 1 g of Sarkosyl in 50 ml DEPC H_2O and filter-sterilize.

CRITICAL: 2% Sarkosyl can be prepared in advance and stored at room temperature for up to a year. CAUTION: Sarkosyl is an irritant and contact with skin, eye or inhalation should be avoided.

1% Tween-20 Dissolve 1 ml of Tween-20 in 49 ml DEPC H_2O and filter-sterilize.

CRITICAL: 1% Tween-20 can be prepared in advance and stored at room temperature for up to a year.

1M Tris-HCl, pH 6.8 Dissolve 6.06 g TRIS base in 50 ml DEPC H_2O, pH to 6.8 with HCl then autoclave and filter-sterilize.

CRITICAL: The buffer can be prepared in advance and stored at room temperature for up to a year.

1M Tris-HCl, pH 7.4 Dissolve 6.06 g TRIS base in 50 ml DEPC H_2O, pH to 7.4 with HCl then autoclave and filter-sterilize.

CRITICAL: The buffer can be prepared in advance and stored at room temperature for up to a year.

1M Tris-HCl, pH 8.0 Dissolve 6.06 g TRIS base in 50 ml DEPC H_2O, pH to 8.0 with HCl then autoclave and filter-sterilize.

CRITICAL: The buffer can be prepared in advance and stored at room temperature for up to a year.

1M DTT Dissolve 1.54 g DTT in 10 ml DEPC H_2O and filter-sterilize.

CRITICAL: 1M DTT can be prepared in advance and stored at -20^{0}C for up to a year.

1mM Biotin-11-CTP Mix 10 μl of 10 mM stock in 90 μl DEPC H_2O to get 1 mM dilution.

CRITICAL: The buffer can be prepared in advance and stored at 4^{0}C for up to a year.

1mM Biotin-11-UTP Mix 10 μl of 10 mM stock in 90 μl DEPC H_2O to get 1 mM dilution.

CRITICAL: The buffer can be prepared in advance and stored at 4^{0}C for up to a year.

50x TAE Dissolve 121 g TRIS base, 28.55 g Glacial Acetic acid, and 50 ml 0.5M EDTA in DEPC H2O to final volume of 500 ml then autoclave and filter-sterilize.

CRITICAL: The buffer can be prepared in advance and stored at 4^{0}C for up to a month.

10x TBE Dissolve 54 g TRIS base, 27.5 g Boric acid, and 20 ml 0.5M EDTA in DEPC H_2O to final volume of 500 ml then autoclave and filter-sterilize.

CRITICAL: The buffer can be prepared in advance and stored at 4^0C for up to a month.

Douncing buffer (for nuclei isolation) 10 mM Tris-HCl pH 7.4, 300mM sucrose, 3 mM $CaCl_2$, 2 mM $MgCl_2$, 0.1% Triton X-100, 0.5 mM DTT, 1 tablet of protease inhibitors cocktail per 50ml, 4 u/ml RNase inhibitor.

CRITICAL: The buffer without DTT, protease inhibitors, and RNase inhibitor can be prepared and stored at 4^0C for up to a month. Add fresh DTT, protease inhibitors, and RNase inhibitor immediately before use.

Permeabilization buffer (for non-yeast cells) 10mM Tris-HCl pH 7.4, 300mM sucrose, 10mM KCl, 5mM $MgCl_2$, 1mM EGTA, 0.05% Tween-20, 0.1% Nonidet P40 substitute, 0.5 mM DTT, 1 tablet of protease inhibitors cocktail per 50ml, 4 u/ml RNase inhibitor.

CRITICAL: The buffer without DTT, protease inhibitors, and RNase inhibitor can be prepared and stored at 4^0C for up to a month. Add fresh DTT, protease inhibitors, and RNase inhibitor immediately before use.

Permeabilization buffer (for yeast cells) 0.5% Sarkosyl, 0.5 mM DTT, 1 tablet of protease inhibitors cocktail per 50ml, 4 u/ml RNase inhibitor.

CRITICAL: The buffer without DTT, protease inhibitors, and RNase inhibitor can be prepared and stored at 4^0C for up to a month. Add fresh DTT, protease inhibitors, and RNase inhibitor immediately before use.

Storage buffer 10 mM Tris-HCl pH 8.0, 25%(vol/vol) glycerol, 5 mM $MgCl_2$, 0.1 mM EDTA, 5 mM DTT.

CRITICAL: The buffer without DTT can be prepared and stored at 4^0C for up to a month. Add fresh DTT immediately before use.

2x Nuclear run-on master mix (for non-yeast cells) 10 mM Tris-HCl pH 8.0, 5 mM $MgCl_2$, 300 mM KCl, and 1 mM DTT.

CRITICAL: The buffer without DTT can be prepared and stored at 4^0C for up to a month. Add fresh DTT immediately before use.

2x NRO master mix (for yeast cells) 40 mM Tris-HCl, pH 7.7, 400 mM KCl, 64 mM $MgCl_2$, 1 mM DTT.)

CRITICAL: The buffer without DTT can be prepared and stored at 4^0C for up to a month. Add fresh DTT immediately before use.

AES buffer (for yeast cells only) 50 mM NaOAc pH 5.3, 10 mM EDTA, 1% SDS.

High-salt wash buffer 50 mM Tris-HCl pH 7.4, 2 M NaCl, 0.5%(vol/vol) Triton X-100 in DEPC H_2O.

CRITICAL: The buffer can be prepared in advance and stored at 4^0C for up to a month.

Binding buffer 10 mM Tris-HCl pH 7.4, 300 mM NaCl, 0.1%(vol/vol) Triton X-100 in DEPC H_2O.

CRITICAL: The buffer can be prepared in advance and stored at 4^0C for up to a month.

Low-salt wash buffer 5 mM Tris-HCl pH 7.4, 0.1%(vol/vol) Triton X-100 in DEPC H_2O.

CRITICAL: The buffer can be prepared and stored at 4^0C for up to a month.

Pre-washed streptavidin-coated magnetic beads Take 90 µl of Streptavidin M280 beads per library. Place the beads on the magnetic separator for 1 min and discard the supernatant. Pre-wash by resuspending in 0.1 N NaOH + 50 mM NaCl in DEPC H_2O for 1 min, place on the magnetic separator for 1 min, remove supernatant. Wash beads twice with 100 mM NaCl in DEPC H_2O. After removing the wash buffer, resuspend the beads in 150 µl of the Binding Buffer and make 3 aliquots of 50 µl each. Scale up accordingly when processing multiple samples.

CRITICAL: The washed beads can be prepared in advance and stored at 4^0C for up to a week.

2.2% Agarose gel 3.3 grams DNA grade agarose in 150 ml 1x TAE. Mix by swirling and heat using a microwave until the mix bubbles and clears.

Gel elution buffer 10 mM Tris-HCl pH 8.0, 0.5 mM NH_4Ac, 10 mM $MgAc_2$, 1 mM EDTA, 0.1% SDS.

CRITICAL: The buffer can be prepared in advance and stored at room temperature for up to a month.

Ammonium persulfate Dissolve 1 g of APS in 10 ml of DEPC-H2O and filter-sterilize the solution with a 0.22-µm filter.

CRITICAL: 10% (wt/vol) APS can be prepared in advance and stored at -20 °C

for up to a year.

DNA and RNA oligos Oligos for PRO-seq and PRO-cap (Table A.1) should be dissolved in DEPC H_2O at a concentration of 100 mM. PCR primers should be dissolved in DEPC H_2O at a concentration of 25 mM.

Software for data analysis

- Adaptor removal software, such as 'cutadapt'(Marcel, 2011) (http://code.google.com/p/cutadapt/)
- Mapping or alignment software, such as 'bwa'(Li and Durbin, 2009) (https://sourceforge.net/projects/bio-bwa/files/) or 'bowtie'(Langmead et al., 2009) (https://sourceforge.net/projects/bowtie-bio/files/)
- Tools for generation of coverage information, such as SAMtools(Li et al., 2009) (https://sourceforge.net/projects/samtools/files/) and BEDTools(Quinlan and Hall, 2010) (https://sourceforge.net/projects/bedtools/files/)

Table A.1. Oligonucleotides required for PRO-seq and PRO-cap
CRITICAL: The DNA & RNA oligos can be stored at -80^{0}C for up to 10 years.

	Oligo name	Sequence (5' to 3')	Purpose	Comments
Oligos for PRO-seq	VRA3	GAUCGUCGGACUG UAGAACUCUGAAC-/Inverted dT/	RNA adaptor for ligation to the 3' end of nascent RNA at step 49	The 5' end is phosphorylated and the 3' end is protected by an inverted dT
	VRA5	CCUUGGCACCCGA GAAUUCCA	RNA adaptor for ligation to the 5' end of nascent RNA at step 75	The 5' end is not phosphorylated
	RP1	AATGATACGGCGA CCACCGAGATCTA CACGTTCAGAGTTC TACAGTCCGA	DNA oligo for reverse transcription of adaptor-ligated nascent RNA at step 84	
Oligos for PRO-cap	RA3	UGGAAUUCUCGGG UGCCAAGG-/Inverted dT/	RNA adaptor for ligation to the 3' end of nascent RNA at step 49	The 5' end is phosphorylated and the 3' end is protected by an inverted dT
	RA5	GUUCAGAGUUCUA CAGUCCGACGAUC	RNA adaptor for ligation to the 5' end of nascent RNA at step 75	The 5' end is not phosphorylated
	RTP	GCCTTGGCACCCG AGAATTCCA	DNA oligo for reverse transcription of adaptor-ligated nascent RNA at step 84	
PCR primers for library amplification	RP1	AATGATACGGCGA CCACCGAGATCTA CACGTTCAGAGTTC TACAGTCCGA	DNA oligo for PCR amplification of cDNA in both PRO-seq and PRO-cap at steps 94 & 102	Same as the RT oligo for PRO-seq
	RPI-n	CAAGCAGAAGACG GCATACGAGAT NNNNNN GTGACTGGAGTT CCTTGGCACCCGA GAATTCCA	DNA oligo with barcodes for PCR amplification of cDNA in both PRO-seq and PRO-cap at steps 94 & 104	The six Ns represent the barcodes for Illumina TRU-seq multiplexing. For example, the barcode in RPI-1 is CGTGAT

Procedure

Cell culture l TIMING 24 h

1. Seed cells at a concentration that will enable them to reach ~80% confluency in 24 hours. For a PRO-seq experiment using adherent fibroblasts, 4-6 150mm cell culture dishes yield sufficient cells (~10^7 cells, see Experimental Design for further details). For yeast cells, plate them to ensure they are in the exponential phase of growth (OD_{600} = 0.5) at the time of harvest.

   CAUTION: Check cell lines for mycoplasma contamination before setting up the experiment.

Sample preparation l TIMING: 1 h

CRITICAL: Samples should be prepared in cold room (4^0C) to avoid unsolicited run-on.

2. Prepare samples by isolating nuclei (Option A) or by cell permeabilization (use Option C for yeast cells and Option B for other cell types).

   All centrifugation steps for sample preparation are performed in a cold centrifuge (4^0C) at 1000 g (unless stated otherwise) for 5 min.

   A. Nuclei isolation

   i. Harvest adherent cells by scraping and centrifuging, and non-adherent cells by centrifuging.

   ii. Resuspend the cell pellet in 10 ml ice-cold PBS and centrifuge.

iii. Resuspend the cell pellet in ice-cold douncing buffer ($1x10^6$ cells/ml).

CRITICAL STEP: If using spike-in cells, add them at this point to the cells resuspended in douncing buffer.

iv. Incubate for 5 min on ice and dounce 25 times using a dounce homogenizer.

v. Transfer the dounced nuclei to either a 15 or 50 ml conical tube and centrifuge the nuclei.

vi. Wash twice by resuspending the pellet in 5 ml douncing buffer and centrifuging.

vii. Resuspend the pellet in storage buffer ($5\text{-}10x10^6$ nuclei per 100 μl of storage buffer), flash freeze in liquid nitrogen, and store at -80^0C.

PAUSE POINT: The nuclei in storage buffer can be stored at -80^0C for up to 5 years.

B. Cell permeabilization (non-yeast cells)

i. Harvest adherent cells by scraping and centrifuging, and non-adherent cells by centrifuging.

ii. Resuspend the cell pellet in 10 ml ice-cold PBS and centrifuge.

iii. Resuspend the cell pellet in ice-cold permeabilization buffer ($1x10^6$ cells/ml).

CRITICAL STEP: Spike-in cells, if used, should be added to the cells resuspended in permeabilization buffer.

iv. Incubate for 5 min on ice and centrifuge the permeabilized cells.

v. Wash twice in 5 ml permeabilization buffer and centrifuging.

vi. Resuspend the cell pellet in storage buffer (5-10x10^6 permeabilized cells per 100 μl of storage buffer), flash freeze in liquid nitrogen, and store in -80^0C.

PAUSE POINT: The permeabilized cells in storage buffer can be stored at -80^0C for up to 5 years.

C. Cell permeabilization (optimized for yeast)

i. Harvest exponentially growing yeast cells by centrifuging at 400 g.

ii. Resuspend the cell pellet in 10 ml ice-cold PBS and centrifuge.

iii. Resuspend the cell pellet in ice-cold yeast permeabilization buffer (1x10^6 cells/ml).

CRITICAL STEP: Spike-in cells, if used, should be added to the cells resuspended in yeast permeabilization buffer.

iv. Incubate for 20 min on ice and centrifuge the cells at 400 g.

v. Resuspend the cell pellet in storage buffer (25-50x10^6 permeabilized cells per 100 μl of storage buffer), flash freeze in liquid nitrogen, and store in -80^0C.

PAUSE POINT: The permeabilized cells in storage buffer can be stored at -80^0C for up to 5 years.

Nuclear run-on | TIMING: 1.5 h

3. Prepare a 2x nuclear run-on (NRO) master mix; for non-yeast cells, prepare the master mix according to the first table, for yeast cells, use the second table.

Reagents	Volume per reaction (μl)	Final concentration – 1x (in 200-μl reaction) (mM)
Tris-Cl pH 8.0 (1 M)	1	5
$MgCl_2$ (1 M)	0.5	2.5
DTT (0.1 M)	1	0.5
KCl (4 M)	7.5	150
DEPC-H_2O	18	

Reagents	Volume per reaction (μl)	Final concentration – 1x (in 200-μl reaction) (mM)
Tris-Cl pH 7.7 (1 M)	4	20
$MgCl_2$ (1 M)	6.4	32
DTT (0.1 M)	1	0.5
KCl (4 M)	10	200
DEPC-H_2O	6.6	

4. Depending on the type of run-on experiment (see Experimental Design section), prepare a 2x reaction mix according to Option A (single biotin run-on), Option B (4 biotin run-on), Option C (2 biotin run-on) or Option D (1 biotin run-on). If processing multiple libraries at once, scale up accordingly (Supplementary Table 1).

A. Individual-biotin run-on 2x reaction mix

i. Transfer a 28 μl aliquot of NRO master mix to each of 4 separate microcentrifuge tubes.

ii. Add 5 μl of biotin-11-ATP (1 mM) to one of the tubes containing NRO master mix. Label this mix “A”

iii. Repeat step ii for the remaining 3 biotin-11-NTPs (1 mM each) and the 3 remaining tubes containing NRO master mix and label them “C”, “G”, “U” accordingly.

iv. Add 15 μl DEPC H_2O to all 4 tubes.

v. Add 2 μl of RNase inhibitor and 50 μl of 2% Sarkosyl to all 4 tubes. From step 5, each tube will be processed as a separate sample.

B. 4-Biotin run-on 2x reaction mix

i. Transfer a 28 μl aliquot of NRO master mix to a microcentrifuge tube.

ii. Add 5 μl each of all 4 biotin-11-NTPs (1 mM each) to the NRO master mix aliquot.

iii. Add 2 μl of RNase inhibitor and 50 μl of 2% Sarkosyl.

C. 2-Biotin run-on 2x reaction mix

i. Transfer a 28 μl aliquot of NRO master mix to a microcentrifuge tube.

ii. Add 5 μl each of biotin-11-CTP (1 mM) and biotin-11-UTP (1 mM) to the NRO master mix aliquot.

iii. Add 2.5 μl each of ATP (10 mM) and GTP (10 mM) to the mix.

iv. Add 5 μl DEPC H_2O.

v. Add 2 μl of RNase inhibitor and 50 μl of 2% Sarkosyl.

D. 1-Biotin run-on 2x reaction mix

i. Transfer a 28 μl aliquot of NRO master mix to a microcentrifuge tube.

ii. Add 5 μl of biotin-11-CTP (1 mM) to the NRO master mix aliquot.

iii. Add 1 μl of CTP (0.05 mM) to the mix.

iv. Add 2.5 μl each of ATP (10 mM), GTP (10 mM), and UTP (10 mM) to the mix.

v. Add 6.5 μl DEPC H_2O.

vi. Add 2 μl of RNase inhibitor and pipette up and down several times.

vii. Add 50 μl of 2% Sarkosyl and pipette up and down 15 times.

5. Preheat 100 μl of the appropriate 2x reaction mix prepared in step 4 to 37°C for mammalian cells or 30°C for yeast and insect cells.

6. Using a cut-off P200 pipette tip, add 100 μl nuclei or permeabilized cells (in storage buffer from step 2) to 100 μl of preheated 2x reaction mix, gently but thoroughly pipette the reaction 15 times, and place in a heat block at the appropriate temperature.

   CRITICAL STEP: Sarkosyl in the 2x reaction mix causes the run-on reaction to become viscous (except for yeast). When adding the nuclei or permeabilized cells to the reaction mix and when mixing by pipetting up and down, use a wide bore pipette tip or cut the last cm off a normal one with ethanol wiped clean scissors or razor blade.

7. Incubate for 3 min (5 min for yeast cells), with gentle tapping at the incubation midpoint.

RNA extraction l TIMING: 1 h

8. Extract RNA using Option A for non-yeast nuclei or permeabilized cells or Option B for yeast.

A. RNA extraction from non-yeast nuclei or permeabilized cells

i. Add 500 μl Trizol LS and mix well by vortexing to stop the reaction.

ii. Incubate the homogenized sample for 5 min at room temperature (25^0C) to allow the complete dissociation of nucleoprotein complexes and add 130 μl Chloroform.

iii. Vortex sample vigorously for 15 s and incubate at room temperature for 2 to 3 min.

iv. Centrifuge at 14,000 *g* for 5 min at 4°C, transfer the aqueous phase to a new tube, and add 1 μl GlycoBlue.

B. RNA extraction from yeast cells or nuclei

i. Pellet cells or nuclei after the run-on reaction at 400 g for 5 mins at 4 °C and quickly resuspend in 500 μl phenol.

ii. Add an equal volume of AES buffer and incubate it at 65°C for 5 min with periodic vortexing. Let the mixture rest on ice for 5 min, and then add 200 μl of chloroform.

iii. Vortex sample vigorously for 15 s and incubate at room temperature for 2 to 3 min.

iv. Centrifuge at 14,000 *g* for 5 min at 4°C, transfer the aqueous phase to a new tube, and add 1 μl GlycoBlue and NaOAc to 200 mM final conc.

9. Add 2.5x volume of 100% room temperature ethanol & vortex for 10 s.
10. Incubate samples at room temperature for 10 min.
11. Centrifuge at 14,000 *g* for 20 min at 4°C. The RNA precipitate forms a gel-like pellet on the side and bottom of the tube.
12. Remove supernatant completely.
13. Add 750 μl of 75% ethanol.

    PAUSE POINT: The RNA pellet in 75% ethanol can be stored up to a week at -80°C.
14. Mix by vortexing and centrifuge at 14,000 *g* for 5 min at 4°C.
15. Remove all supernatant.
16. Air-dry the RNA pellet for 5-10 min.

    CRITICAL STEP: It is important not to let the RNA pellet dry completely as this will greatly decrease its solubility.
17. For PRO-seq, re-dissolve the RNA pellet in 20 μl DEPC H_2O and proceed to step 18. For PRO-cap, re-dissolve the RNA pellet in 50 μl DEPC H_2O and proceed to step 23.

RNA fragmentation by base hydrolysis (PRO-seq only) | Timing: 0.5 h

18. Heat denature the RNA at 65°C on a heat block for 40 s and then place the tubes on ice.
19. Add 5 μl of cold 1 N NaOH and incubate the mixture on ice for 10 min.
20. Add 25 μl of 1 M Tris-HCl pH 6.8.

21. Perform buffer exchange once by running the 50-μl base-hydrolyzed RNA through a P-30 column following manufacturer's instructions.
22. Add 1μl RNase inhibitor.

Biotin RNA enrichment | Timing: 3 h

23. Mix ~50 μl of the RNA sample from step 22 for PRO-seq or from step 17 for PRO-cap with 50 μl of pre-washed Streptavidin beads.
24. Incubate at room temperature on a rotator set at 8 rpm for 20 min.
25. Place on magnet for 1 min and remove the liquid.
26. Resuspend the beads in 500 μl ice cold High salt wash buffer for a 1 min wash using rotator.
27. Place on magnet for 1 min and remove the buffer.
28. Repeat steps 26-27 once more.
29. Wash two times with 500 μl Binding buffer for a minute and use the magnet to facilitate removal of buffer.
30. Wash once with 500 μl Low salt wash buffer.
31. Resuspend the beads in 300 μl Trizol and vortex thoroughly.
32. Incubate for 3 min at room temperature.
33. Add 60 μl chloroform.

    CRITICAL STEP: Inaccurate pipetting of chloroform leads to incomplete phase separation of the Trizol when transferring a small volume.

34. Vortex thoroughly for more than 20 s and incubate the tubes for 3 min at room temperature.
35. Centrifuge at 14,000 *g* for 5 min, 4°C.
36. Transfer ~180 µl of the aqueous layer to a new tube.
37. Remove and discard the organic phase, leaving the beads and the unpipetted aqueous phase.
38. Extract RNA from the beads once more by repeating steps 31-35.
39. Collect ~180 µl of the aqueous layer and combine with the sample from step 36.
40. Add 360 µl chloroform to the pooled aqueous layer and vortex.
41. Centrifuge at 14,000 *g* for 5 min, 4°C.
42. Transfer ~350 µl of the aqueous layer to a clean tube.
43. To the collected aqueous layer, add 1 µl of GlycoBlue, 900 µl of 100% ethanol and vortex.
44. Incubate samples at room temperature for 10 min and centrifuge at 14,000 *g* for 20 min, 4°C.

    TROUBLESHOOTING:
45. Add 750 µl of 75% ethanol.

    PAUSE POINT: The RNA pellet in 75% ethanol can be stored up to a week at -80°C.
46. Mix by vortexing and centrifuge at 14,000 *g* for 5 min, 4°C.
47. Remove all residual liquid.

48. Air-dry the RNA pellet for 5-10 min.

    CRITICAL STEP: Do not re-dissolve in H_2O without the RNA adaptor. The RNA pellet is dissolved in small volume of RNA adaptor-containing solution to minimize the adapter ligation reaction volume.

3' RNA adaptor ligation I Timing: 4.5 h

49. Dilute 0.5 µl 100 µM 3' RNA adaptor in 3.5 µl DEPC H_2O. For PRO-seq, use VRA3 RNA adaptor. For PRO-cap, use RA3 RNA adaptor. For processing multiple samples, scale up accordingly.
50. Redissolve the RNA from step 48 in 4 µl of 3' RNA adaptor dilution.
51. Heat denature at 65°C in a heat block for 20 s, then place on ice.
52. Make the RNA ligation mix tabulated below. When processing multiple samples, scale up accordingly (Supplementary Table 2).

Reagents	**Volume per reaction (µl)**	**Final concentration**
T4 RNA ligase buffer (10x)	1	1x
ATP (10 mM)	1	1 mM
50% PEG	2	10 %
RNase inhibitor	1	4 units per µl
T4 RNA ligase I	1	1 units per µl

53. Add 6 µl of the mix to the 4 µl of RNA (10 µl final).

    CRITICAL STEP: GlycoBlue may form a precipitate in presence of high PEG conc, but is not reported to affect the ligation efficiency.

54. Incubate at 20°C for 4 hr then place at 4°C until ready to proceed to the next step.

    PAUSE POINT: The ligation reaction can be left at 4°C overnight.

Second biotin RNA enrichment | Timing: 3 h

55. Bring the volume of the adaptor ligated RNA from step 54 to 50 µl by adding 40 µl DEPC H_2O.
56. Perform a second biotin enrichment by repeating steps 23-48 with the 50 µl ligated RNA sample.

    PAUSE POINT: The RNA pellet in 75% ethanol can be stored up to a week at -80°C.

Enzymatic modification of the RNA 5' ends | Timing: 3.5-4 h

57. Re-dissolve the RNA pellet from step 56 in 5 µl DEPC H_2O.
58. Heat denature briefly at 65°C in a heat block for 20 secs, then place on ice.
59. If performing PRO-cap, degrade the uncapped RNA containing 5'-monophosphate and remove the 5' triphosphate and monophosphate from uncapped RNA as described in Box 1 (Supplementary Table 3 & 4) before proceeding to step 60. For PRO-seq, continue directly to step 60.
60. Prepare 5' cap repair enzyme mix: depending on the availability of Tobacco acid pyrophosphatase (TAP) or RNA 5' Pyrophosphohydrolase

(RppH). When using TAP, prepare the enzyme mix in the first table. When using RppH, prepare the enzyme mix in the second table. When processing multiple samples, scale up (Supplementary Table 5).

Reagents	Volume per reaction (μl)	Final concentration
DEPC H_2O	3	
TAP buffer (10x)	1	1x
RNase inhibitor	0.5	2 units per μl
TAP	0.5	0.5 units per μl

Reagents	Volume per reaction (μl)	Final concentration
DEPC H_2O	2.5	
ThermoPol Reaction buffer (10x)	1	1x
RNase inhibitor	0.5	2 units per μl
RppH	1	0.5 units per μl

61. Add 5μl of the appropriate enzyme mix to the 5μl of RNA from step 58 for PRO-seq or 5μl of RNA from step 24 of Box 1 for PRO-cap.

62. Incubate at 37°C for 1 hr.

63. For PRO-seq, proceed to step 64 for hydroxyl repair. For PRO-cap, Add 90μl of DEPC H_2O to the 10μl of RNA from step 62 and proceed directly to step 67

(PRO-seq only) Hydroxyl repair l TIMING: 2 h

64. Prepare polynucleotide kinase (PNK) mix as tabulated below. When processing multiple samples, scale up accordingly (Supplementary Table 6).

Reagents	Volume per reaction (μl)	Final concentration
DEPC H_2O	65	
PNK buffer (10x)	10	1x
ATP (10 mM)	10	1 mM
RNase inhibitor	2.5	1 units per μl
PNK	2.5	0.25 units per μl

65. Add 90μl of the mix to the 10μl of RNA from step 62.
66. Incubate at 37°C for 1 hr.
67. Add 300 μl of Trizol and vortex for 5 s.
68. Add 60 μl chloroform, vortex for 15 s, and incubate for 2 min at room temperature.
69. Centrifuge at 14,000 *g* for 5 min at 4°C.
70. Transfer ~280 μl aqueous layer to a clean microfuge tube.
71. Add 280 μl chloroform to the aqueous layer from step 70 and vortex for 5 s.
72. Centrifuge at 14,000 *g* for 5 min at 4°C and transfer ~280 μl aqueous layer to a clean microfuge tube.
73. Add 0.5 μl GlycoBlue and 700 μl of 100% ethanol to the aqueous layer from step 72, and pellet the RNA by centrifuging at 14,000 *g* for 20 min at 4°C.
74. Wash the RNA pellet in 75% ethanol by repeating steps 45-48.

    PAUSE POINT: The RNA pellet in 75% ethanol can be stored up to a week at -80°C.

CRITICAL STEP: Do not re-dissolve in H_2O without the RNA adaptor. The RNA pellet is dissolved in a small volume of RNA adaptor-containing solution to minimize the adapter ligation reaction volume.

5' RNA adaptor ligation l Timing: 4.5 h

75. Dilute 0.5 µl 100 µM 5' RNA adaptor in 3.5 µl DEPC H_2O. For PRO-seq, use VRA5 RNA adaptor. For PRO-cap, use RA5 RNA adaptor. For processing multiple samples, scale up accordingly.
76. Redissolve the RNA pellet from step 74 in 4 µl of the 5' RNA adaptor dilution.
77. Heat denature at 65°C in a heat block for 20 s, then place on ice.
78. Make the RNA ligation mix as described in step 52. When processing multiple samples, scale up accordingly (Supplementary Table 2).
79. Add 6 µl of the RNA ligation mix to the 4 µl of RNA (10 µl final).
    CRITICAL STEP: GlycoBlue may precipitate in the presence of high PEG concentration, but is not reported to affect the ligation efficiency.
80. Incubate at 20°C for 4 hr then place at 4°C until ready to proceed to the next step.
    PAUSE POINT: The ligation reaction can be left at 4°C overnight.

Third biotin RNA enrichment | Timing: 3 hr

81. Bring the volume of the adaptor ligated RNA to 50 µl by adding 40 µl DEPC H_2O.

82. Perform a third biotin enrichment by repeating steps 23-48 with the 50 µl ligated RNA sample.

    PAUSE POINT: The RNA pellet in 75% ethanol can be stored up to a week at -80°C.

Reverse transcription | Timing: 2 h

83. Re-dissolve the RNA pellet in 10 µl DEPC H_2O.

84. Make reverse transcription (RT) primer mix as shown in the table below.

Component	Amount per reaction (µl)		Final conc. in 20-µl volume (µM)
	PRO-seq	PRO-cap	
RP1 reverse transcription primer (100 µM)	0.5	--	2.5
RTP reverse transcription primer (100 µM)	--	0.5	2.5
12.5 mM dNTP mix	1	1	625
DEPC-H_2O	1	1	

85. Add 2.5 µl of the RT primer mix to the 10 µl of re-dissolved RNA.

86. Heat to 70°C for 2 min, chill it on ice for 2 min and briefly spin at 500-1,000g at 25°C for 5 s.

87. Prepare the RT buffer mix as shown in table below. When processing multiple samples, scale up accordingly (Supplementary Table 7).

Reagents	Volume per reaction (μl)	Final concentration
First strand buffer (5x)	4	
DTT (0.1 M)	1	5 mM
RNase inhibitor	1	2 units per μl

88. Add 6 μl of RT buffer mix to 12.5 ul of RNA-primer mix from step 86.

89. Incubate for 5 min, 37°C.

90. Add 1.5 μl Superscript III reverse transcriptase and mix (total 20 μl).

91. Incubate for 15 min at 45°C, then 40 min at 50°C, 10 min at 55°C, and 15 min at 70°C.

92. Add 6 μl of DEPC H_2O to the RT reaction (total 26 μl).

    PAUSE POINT: The reverse transcribed cDNA can be stored for a month at -20°C.

Test PCR amplification l Timing: 2 h

93. Prepare a series of 4-fold dilutions of the RT sample in H_2O as tabulated below. Test PCR amplification (a total of 21 amplification cycles) of these dilutions will be used to determine the appropriate number of PCR cycles to use in full-scale amplification at step 105. The use of 2 μl RT sample in dilution 1 leaves 24 μl for full-scale amplification at step 105.

Dilution	Amount of cDNA	Amount of H_2O (μl)	Equivalent full-scale PCR cycles (step 105)
1	2 μl RT sample	6	17
2	2 μl dilution 1	6	15
3	2 μl dilution 2	6	13
4	2 μl dilution 3	6	11

CRITICAL STEP: As only 6 µl out of the total 8 µl of Dilution 1 is used for PCR, this is equivalent to using 1.5 µl (6/8 x 2) of the original 26 µl RT sample, which is 16 fold less than the remaining 24 µl RT sample. Thus, to account for the 16-fold higher amount of starting material in the full-scale amplification, the number of PCR cycles needs to be reduced by 4 compared to test PCR of Dilution 1 i.e. 17 cycles instead of 21. Each remaining dilution in the 4-fold dilution series will need to be corrected by a further 2 cycles compared to the previous dilution i.e. 15, 13, and 11 cycles for dilutions 2 to 4 respectively.

94. Prepare a test PCR mix. One test PCR amplification will be performed for each dilution prepared in step 93 (Supplementary Table 8).

Reagents	**Volume per reaction (µl)**	**Final concentration**
DEPC H_2O	5	
HF buffer (5x)	4	1x
Betaine (5 M)	4	1 M
dNTP mix (12.5 mM each)	0.4	250 µM each
RP1 primer (25 µM)	0.2	250 nM
RPI-1 primer (25 µM)	0.2	250 nM
Phusion DNA polymerase	0.2	0.02 units per µl

95. Add 14 µl of the PCR mix to 6 µl of each diluted test sample.

96. Use the following thermal cycling to perform test PCR amplification (a total of 21 amplification cycles).

Cycle number	**Denature (95°C)**	**Anneal**	**Extend (72°C)**
1	2 min		
2 - 6	30 s	56°C for 30 s	30 s

7 - 22	30 s	65°C for 30 s	30 s
23			10 min

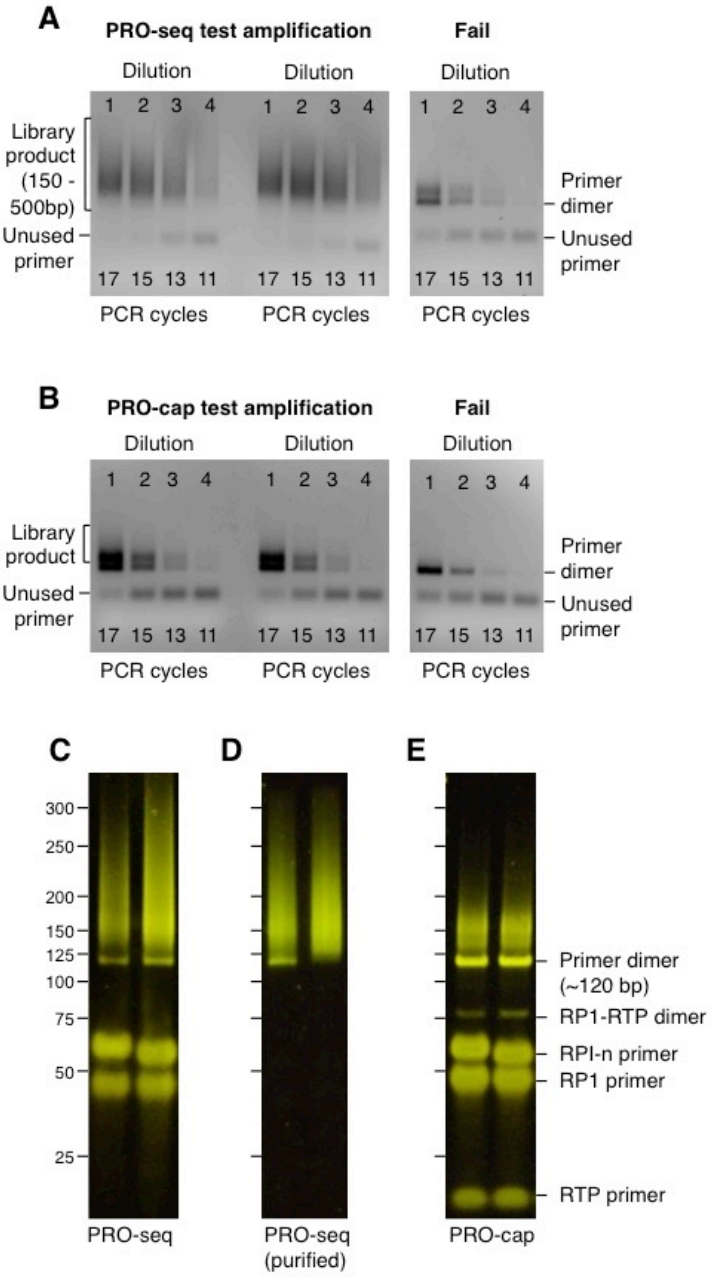
Gel analysis of test PCR I Timing: 2 h

97. Add 2.2 µl 10x Orange G dye to the 20 µl PCR reactions and load 20 µl of the samples onto a 2.2% agarose gel in 1xTAE.
98. Load 8 µl of 100 bp DNA ladder on a separate lane.
99. Run the gel at 100 V for 15 min, then run it at 130 V for up to 45 min.

    CRITICAL STEP: Orange G dye runs at 50 bp. Stop the electrophoresis before the dye runs out of the gel.
100. Add 15 µl of SYBRGold to 150 ml 1x TAE. Place the gel in this solution and stain for 30 min on a shaker.
101. Image the gel with 485 nm illumination, or with UV light. Examine the gel and determine the dilution (and therefore the equivalent full-scale PCR amplification cycle) with desired amplification characteristics (sufficient amount of product, not over amplified, and having 50-75% of unused primers). For example, if the lane in the agarose gel with dilution 3 has the desired amplification characteristics, then the optimal number of PCR cycles (optimized cycle, OC) for full-scale amplification is 13 (see step 93). See Figure A.2A and A.2B for an example gel image.

    TROUBLESHOOTING:

Figure A.2. Gel images of library products
(A) PRO-seq libraries at the test amplification stage (step 101). 3 different samples are loaded on the agarose gel, each at 4 dilutions. From the left of each series, the 4 dilutions are equivalent to 17, 15, 13, and 11 cycles of full amplification. The gel image in left panel shows two successful PRO-seq libraries. Optimal amplification cycles are determined by comparing both the intensities of library products and the unused primers. Optimal amplification cycles are when 50-75% of the primers remain unused (13 cycles for this example). Gel image of a failed library preparation is shown on the right panel; only primer dimers are detected.
(B) PRO-cap libraries at the test amplification stage (step 101). Library products are smaller in size and amount, because most PRO-cap RNA molecules are from paused RNA polymerase. 15 cycles were optimal for these samples.
(C) Two PRO-seq libraries after full amplification analyzed by 8% PAGE (step 118). Labels on the left indicate DNA sizes. Note the presence of unused primers around 50 bp and primer dimer bands ~120 bp. The library product is the smear above the primer dimer band.
(D) Two PRO-seq libraries after the size selection analyzed by 8% PAGE (step 119). PRO-seq library on the left still has some residual primer dimer.
(E) Two PRO-cap libraries after the full amplification analyzed by 8% PAGE (step 118). Note that there are additional primer and primer dimer bands compared to PRO-seq.


A
PRO-seq test amplification
Fail
Dilution
Dilution
Dilution
1 2 3 4
1 2 3 4
1 2 3 4
Library product (150 - 500bp)
Unused primer
Primer dimer
Unused primer
17 15 13 11
17 15 13 11
17 15 13 11
PCR cycles
PCR cycles
PCR cycles
B
PRO-cap test amplification
Fail
Dilution
Dilution
Dilution
1 2 3 4
1 2 3 4
1 2 3 4
Library product
Unused primer
Primer dimer
Unused primer
17 15 13 11
17 15 13 11
17 15 13 11
PCR cycles
PCR cycles
PCR cycles
C
D
E
300
250
200
150
125
100
75
50
25
Primer dimer (~120 bp)
RP1-RTP dimer
RPI-n primer
RP1 primer
RTP primer
PRO-seq
PRO-seq (purified)
PRO-cap

Full-scale PCR amplification I Timing: 2.5 h

102. Prepare a full-scale amplification PCR mix according to the table below. For processing multiple samples, scale up accordingly (Supplementary Table 9).

Reagents	Volume per reaction (μl)	Final concentration
DEPC H_2O	3	
HF buffer (5x)	10	1x
Betaine (5 M)	10	1 M
dNTP mix (12.5 mM each)	1	250 μM each
RP1 primer (25 μM)	0.5	250 nM
Phusion DNA polymerase	1	0.04 units per μl

103. Add 25.5 μl of the PCR mix to the 24 μl of the remaining reverse transcription product (from step 93).

104. Add 0.5 μl of a different RPI-n (25 μM) primer to the different libraries so that each library is differentially barcoded.

CRITICAL STEP Different barcodes are only needed if the libraries are to be pooled for sequencing.

105. Use the following thermal cycling for full-scale PCR amplification, where OC stands for Optimized Cycle as determined in step 101.

Cycle number	Denature (95°C)	Anneal	Extend (72°C)
1	2 min		
2 - 6	30 s	56°C for 30 s	30 s
7 - (OC + 1)	30 s	65°C for 30 s	30 s
(OC + 1) - (OC + 2)			10 min

PAUSE POINT: The PCR product can be stored up to a month at -20°C.

106. Add 231 µl H_2O, 18 µl 5 M NaCl, 1 µl GlycoBlue to the 50 µl full amplification PCR product.
107. Add 750 µl 100% ethanol and vortex thoroughly.
108. Centrifuge at 14,000 *g* for 20 min, 4°C.
109. Remove liquid and the wash the pellet once in 75% ethanol by repeating steps 45-48.

    PAUSE POINT: The DNA pellet in 75% ethanol can be stored up to several months at -80°C.
110. Re-dissolve the DNA pellet in 18 µl H_2O.

Library size selection by PAGE l Timing: 5 h to 1 day

CRITICAL: we describe PAGE purification for size selection of the library. However, a Pippin Prep (Sage Science) can also be used in place of steps 111-130; follow the manufacturer's instruction for selecting a size range of 140-350 bp.

111. Add 2 µl of 10× Orange G loading dye to 18 µl DNA from step 110.
112. Prepare a medium size (10 cm running length) Native PAGE gel as below.

Reagents	Volume	Final concentration
DEPC-H_2O	31.67 ml	
Acrylamide (30%)	13.3 ml	8%
TBE (5x)	5 ml	0.5x
APS (10%)	500 µl	0.1%
TEMED	50 µl	

113. Pre-run the gel for 15 min at a constant current of 30 mA.
114. Load samples. Also load 2 µl 10 bp DNA ladder and 8 µl of 100 bp DNA ladder.
115. Run gel at 15 mA for 30 min until the DNA has entered the gel, and then run at 30 mA for 1.5 hr. Stop electrophoresis 10 minutes after the Orange G dye has run off the gel.
116. During the electrophoresis, puncture the bottom of a sterile, nuclease-free, 0.5 ml centrifuge tube using a 21-gauge needle (heated in a bunsen flame) to create a hole or several holes in the bottom of the tube. Place the 0.5 ml microtube into a sterile, round-bottom, nuclease-free, 2 ml microtube.
117. After the electrophoresis, pry apart the gel cassette and stain the gel with SYBR Gold (10 µl SybrGold per 100 ml 1× TBE buffer) in a clean container for 5-10 min.
118. Visualize the gel on a Dark Reader transilluminator.
119. Using a clean scalpel or razor, cut the gel from 140 bp (20 bp just above the 120 bp adapter dimer) up to 350 bp (see Figure A.2C, A.2D, and A.2E).
120. Split the gel fragment vertically and place the pieces into the 0.5 ml microtube.

121. Centrifuge the stacked tubes at 10,000 *g* for 2 min at room temperature to shred the gel through the holes into the 2 ml tube (there is no liquid at this point).
122. If some gel remains in the top tube, add 100 µl of Gel Elution Buffer and spin at 10,000g again for another 2 min.
123. Add 600 µl of Gel Elution Buffer and incubate for 2 hr in a rotating incubator, 37°C.

    PAUSE POINT: The elution can continue overnight.
124. Spin down gel pieces for 1 min at max speed in a benchtop centrifuge.
125. Transfer all liquid possible to a new tube.
126. Add 400 µl of Gel Elution Buffer to the remaining gel pieces.
127. Incubate for 1 hr in a rotating incubator, 37°C.
128. After 1 h incubation, spin at the maximum speed in a benchtop centrifuge for 1 min; take the supernatant and pool with the first elution from step 125.
129. Rinse gel pieces with 250 µl H_2O, spin and add the rinsed liquid to the pool.
130. Transfer the pooled eluate which may contain small pieces of gel debris, to the top of a Spin-X filter.
131. Centrifuge the filter for 1-2 min at 6,000-7,000 *g*, room temperature. Collect the filtrate. If the volume exceeds the filter capacity, use multiple

filters or split into batches and repeat filtering a couple of times and pool the filtrates.

132. Lyophilize (on medium setting) the sample using a Speed Vac dryer reduce the volume to ~400 µl (takes 45 min-2 hr). If the volume decreases below 400 µl, bring the volume up to 400 µl by adding DEPC-H_2O.
133. Add an equal volume of buffered phenol:chloroform, and vortex thoroughly.
134. Centrifuge at 14,000 *g* for 5 min, 4°C.
135. Collect the aqueous layer in a clean tube.
136. Add an equal volume of chloroform to the aqueous layer, and vortex thoroughly.
137. Centrifuge at 14,000 *g* for 5 min, 4°C.
138. Collect the aqueous layer in a clean tube.
139. Add 1 µl GlycoBlue to the aqueous layer.
140. Add 2.5× volume of room temperature 100% ethanol.
141. Vortex thoroughly and incubate at room temperature for 10 min.
142. Centrifuge at 14,000 *g* for 20 min, 4°C.
143. Remove liquid, and wash the DNA pellet once in 75% ethanol by repeating steps 45-48.
144. Re-dissolve the pellet in 12 µl H_2O.

145. Use 2 µl of the library DNA for quantification using Qubit or Bioanalyzer. The expected concentration of the library is between 1-20 ng/µl.

146. If required, dilute the samples to 5 ng/µl. Send ~10 ng to the sequencing facility for sequencing. If the libraries are barcoded, pool the barcoded libraries that are to be sequenced simultaneously.

High-throughput sequencing | Timing: 24 h

147. Sequence pooled PRO-seq or PRO-cap libraries using an Illumina TRU-seq compatible sequencing platform. Sequencing depth of ~20 million and ~50 million reads provides good coverage in insect cells and mammalian cells respectively.

Data analysis | TIMING: variable

CRITICAL. In PRO-seq, the 3' end of the nascent RNA corresponds to the genomic position of the RNA polymerase active site. The modified RNA adaptors were designed to enable sequencing of the reverse complement of nascent RNAs. Therefore, the 3' end of the reverse complement of the sequencing reads reflects the RNA polymerase active site position. In PRO-cap, conventional RNA adaptors are used, and the 5' ends of each sequence read reflects the transcription start sites in the same direction. Below, we outline the three major stages of a simple processing pipeline.

148. *Pre-process the raw sequence data*. Filter out low quality reads and

trimming potential adaptor sequences (TGGAATTCTCGGGTGCCAAGG) from the sequence reads. Tools such as ‘cutadapt’ are publicly available for this purpose. Depending on the quality of the library, sequences containing only the adaptor sequences (adaptor dimers) may take up to 5% of total reads.

149. *Map or align the sequence reads to the genomic sequence*. Since most of the nascent RNA reads are captured before RNA processing and splicing, they do not contain large gaps in alignment. Therefore, many alignment programs based on the Burrows-Wheeler transformation algorithm such as ‘bwa’ or ‘bowtie’ work well. Reads with multiple alignments are usually discarded, unless they are used for studying specific target regions that are repeated more than once. Sometimes, reads aligning to the ribosomal DNA sequence can be pre-filtered since they can account for 30-40% of all the transcriptional activity. On average, about 55-70% of the raw sequence reads are aligned uniquely to the genome. The alignment results are commonly stored in ‘sam’ or ‘bam’ formats.

150. *Generate the coverage of the aligned sequence reads*. A common way to do this using publicly available tools is as follows: first, sort the ‘bam’ file using ‘samtools sort’; then process the sorted ‘bam’ file using ‘bedtools genomecov’ with ‘-ibam’ (use bam file input), ‘-strand’ (strand specific coverage), and ‘-5’ (5’ position coverage’) options. For the PRO-

seq data, swap the plus and minus strand data for the correct orientation. These data can be visualized in genome browsers (Figure A.3), and used in further downstream analyses.

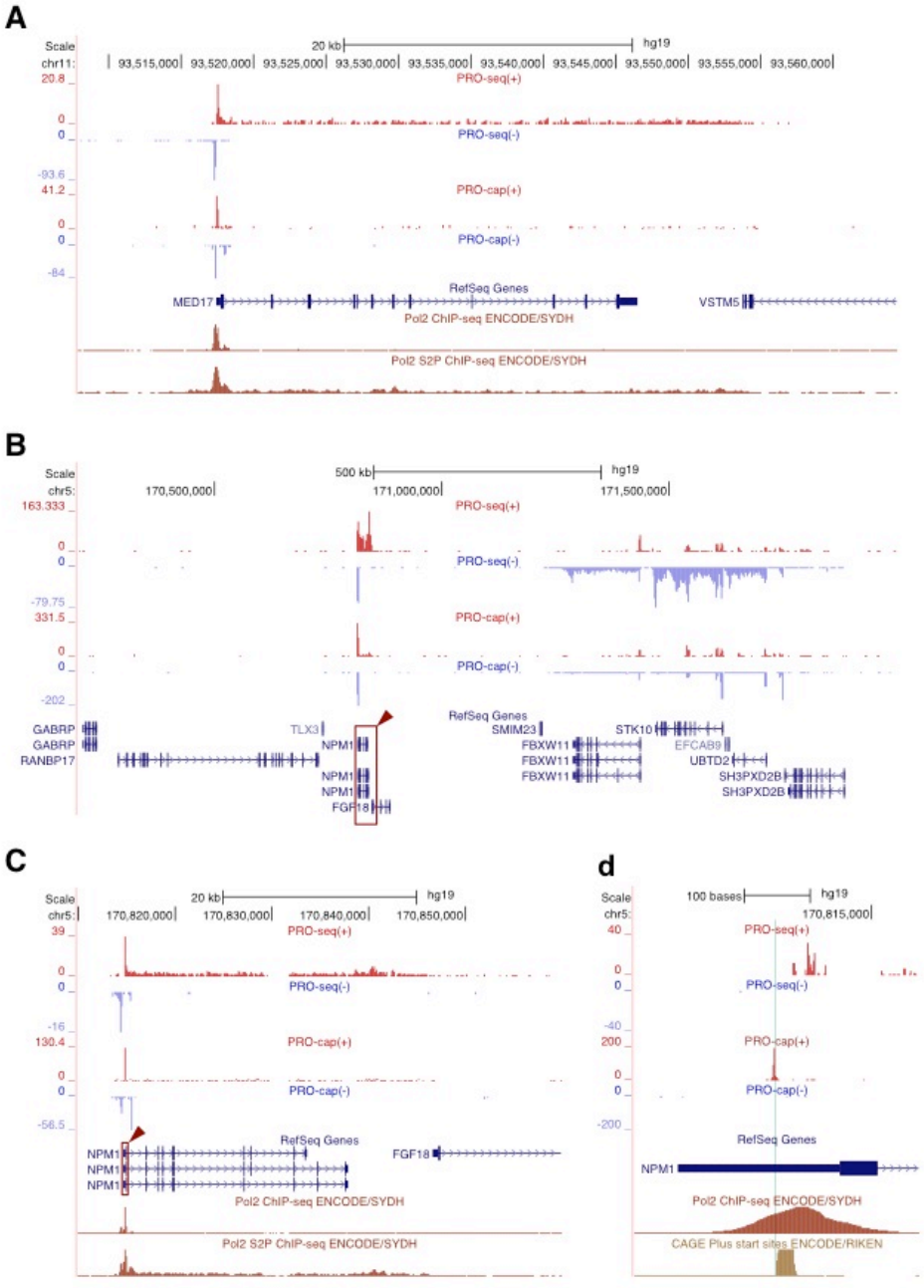
TROUBLESHOOTING

Troubleshooting advice is provided in Table A.2.

Table A.2. Troubleshooting PRO-seq library preparation

Step	Problem	Possible reason	Solution
44	Small or no pellet	Incomplete precipitation.	· Re-centrifuge using higher speed. · If the pellet is still not recovered, add 0.4× volume of isopropanol and re-centrifuge.
		Incomplete separation of the organic phase	· Perform an additional chloroform extraction after step 42.
	Blue jelly ball-like pellet	Incomplete separation of the organic phase.	· Re-dissolve the pellet in 100 µl DEPC H_2O and extract RNA once more using Trizol LS. · If the problem persists, add additional chloroform extraction after step 42.
101	No library product or the library product amplifies at later cycles.	Insufficient amount or quality of the nuclei sample.	· Adjust the nuclei amount and monitor the quality. A typical batch of active nuclei can incorporate 1-5% of total radioactive CTP [α-^{32}P] under the nuclear run-on condition.
		RNA degradation.	· Replace all reagents using RNA-grade materials and carry out steps at low 4^{0}C.
		Incomplete RNA extraction during biotin RNA enrichment.	· Resuspend the bead thoroughly in step 31. · Add additional RNA extraction by repeating step 38.
		Inefficient RNA modifying enzymes.	· Monitor enzyme activities following the manufacturers' instructions.

Figure A.3. Genome browser examples of PRO-seq and PRO-cap results
(A) Sample of PRO-seq and PRO-cap data viewed on the University of California at Santa Cruz (UCSC) genome browser. A region on chromosome 11 encompassing the *MED17* gene showing PRO-seq (top), PRO-cap (mid), and Pol II ChIP-seq and Serine-2 phosphorylated Pol II (Pol II S2P) ChIP-seq (bottom) reads aligned to the human genome (hg19). PRO-seq and PRO-cap reads on the plus strand (left to right), red; PRO-seq and PRO-cap reads on the minus strand (right to left), light blue; RefSeq gene annotations, dark blue; and ChIP-seq reads, dark red. The Pol II and Pol II S2P ChIP-seq tracks are from the ENCODE public data on GM12878 cells using 8WG16 and ab5095 antibodies respectively. Y-axis represents raw read counts displayed on the default setting of the UCSC genome browser.
(B) Sample of PRO-seq and PRO-cap data view on the UCSC genome browser. A region on chromosome 5 encompassing the *NPM1* gene showing PRO-seq (top) and PRO-cap (bottom) reads aligned to the human genome (hg19). Red boxes with arrowheads mark the region magnified in subsequent panels.
(C) PRO-seq (top), PRO-cap (mid), and Pol II ChIP-seq and Pol II S2P ChIP-seq (bottom) data across the gene body of *NPM1* gene.
(D) An up-close view around the annotated TSS of *NPM1* gene. ENCODE RIKEN CAGE data (bottom) is shown to illustrate the position of mRNA cap site relative to the PRO-seq, PRO-cap, and Pol II ChIP-seq data around the annotated TSS of *NPM1* gene.


A
Scale
chr11:
20 kb
hg19
93,515,000
93,520,000
93,525,000
93,530,000
93,535,000
93,540,000
93,545,000
93,550,000
93,555,000
93,560,000
PRO-seq(+)
PRO-seq(-)
PRO-cap(+)
PRO-cap(-)
RefSeq Genes
MED17
VSTM5
Pol2 ChIP-seq ENCODE/SYDH
Pol2 S2P ChIP-seq ENCODE/SYDH
B
Scale
chr5:
500 kb
hg19
170,500,000
171,000,000
171,500,000
PRO-seq(+)
PRO-seq(-)
PRO-cap(+)
PRO-cap(-)
RefSeq Genes
GABRP
GABRP
RANBP17
TLX3
NPM1
FGF18
SMIM23
STK10
FBXW11
EFCAB9
UBTD2
SH3PXD2B
C
Scale
chr5:
20 kb
hg19
170,820,000
170,830,000
170,840,000
170,850,000
PRO-seq(+)
PRO-seq(-)
PRO-cap(+)
PRO-cap(-)
RefSeq Genes
NPM1
FGF18
Pol2 ChIP-seq ENCODE/SYDH
Pol2 S2P ChIP-seq ENCODE/SYDH
d
Scale
chr5:
100 bases
hg19
170,815,000
PRO-seq(+)
PRO-seq(-)
PRO-cap(+)
PRO-cap(-)
RefSeq Genes
NPM1
Pol2 ChIP-seq ENCODE/SYDH
CAGE Plus start sites ENCODE/RIKEN

Anticipated results

The final product of the PRO-seq or PRO-cap method is genome-wide maps of, respectively, RNA polymerase active sites or transcription start sites (Figure A.3). In general, PRO-seq profile of a gene exhibits several features such as: a) higher read density at the promoter-proximal pause site compared to the gene body representing paused Pol IIs, b) uniform read density across exons and introns, c) reads beyond the polyadenylation site, and d) divergent PRO-seq reads at the promoter of genes in mammals indicating divergent transcription. The enhancer regions, which are present in intergenic and intragenic regions, are also characterized by divergent PRO-seq reads. In PRO-cap, read density is very high at TSS and very low, almost at background level, in gene bodies.

An indication of whether library preparation has been successful can be obtained at the Test amplification stage (steps 93-101) through estimates of yield and quality (Figure A.2). Spiky PRO-seq read coverage along the gene body indicates lower complexity of the library, which may arise from use of fewer nuclei or permeabilized cells or by the use of higher number of PCR amplification cycles. Libraries that require fewer amplification cycles (less than 13 cycles) provide high quality results, and those requiring between 14-18 cycles provide meaningful results but with some potential for amplification biases. These amplification biases manifest as low library complexity in general and high repetition of certain sequence reads.

Box 1. Degradation of RNA containing 5'-monophosphate and removal of 5'-triphosphate and -monophosphate from the RNA for PRO-cap I

Timing: 4 h

1. Prepare the 5'-phosphate-dependent exonuclease enzyme mix. When processing multiple samples, scale up accordingly (Supplementary Table 3)

Reagents	Volume per reaction (µl)	Final concentration
DEPC-H_2O	2.5	
Buffer A (10x)	1	1x
RNase inhibitor	0.5	2 units/µl
5'-phosphate-dependent exonuclease	1	0.1 units/µl

2. Add 5 µl of the mix to the RNA from step 58 of the main Procedure (10µl final)
3. Incubate at 30°C for 1 hr.
4. Add 300 µl of Trizol and vortex for 5 s.
5. Add 60 µl chloroform, vortex for 15 s, and incubate for 2 min at room temperature.
6. Centrifuge at 14,000 *g* for 5 min at 4°C.
7. Transfer ~180 µl aqueous layer to a clean microfuge tube.
8. Add 180 µl chloroform to the aqueous layer from step 7 and vortex for 5 s.
9. Centrifuge at 14,000 *g* for 5 min at 4°C and collect ~180 µl aqueous layer.

10. Add 0.5 µl GlycoBlue and 450 µl of 100% ethanol to the aqueous layer from step 9, and pellet the RNA by centrifuging at 14,000 *g* for 20 min at 4°C.
11. Wash the RNA pellet in 75% ethanol by repeating steps 45-48 of the main Procedure.

    PAUSE POINT: The RNA pellet in 75% ethanol can be stored up to a week at -80°C.
12. Re-dissolve the RNA pellet in 5 µl DEPC H_2O, and heat denature briefly in 65°C heat block for 20 s, then place on ice.

13. Prepare the alkaline phosphatase enzyme mix. When processing multiple samples, scale up accordingly (Supplementary Table 4).

Reagents	**Volume per reaction (µl)**	**Final concentration**
DEPC-H_2O	3	
Alkaline phosphatase buffer (10x)	1	1x
RNase inhibitor	0.5	2 units/µl
Alkaline phosphatase	0.5	0.5 units/µl

14. Add 5 µl of the mix to the RNA from step 12 (10µl final)
15. Incubate at 37°C for 1 hr.
16. Add 300 µl of Trizol and vortex for 5 s.
17. Add 60 µl chloroform, vortex for 15 s, and incubate for 2 min at room temperature.

18. Centrifuge at 14,000 *g* for 5 min at 4°C.

19. Transfer ~180 µl aqueous layer to a clean microfuge tube.

20. Add 180 µl chloroform to the aqueous layer from step 19 and vortex for 5 s.

21. Centrifuge at 14,000 *g* for 5 min at 4°C and collect ~180 µl aqueous layer.

22. Add 0.5 µl GlycoBlue and 450 µl of 100% ethanol to the aqueous layer from step 21, and pellet the RNA by centrifuging at 14,000 *g* for 20 min at 4°C.

23. Wash the RNA pellet in 75% ethanol by repeating steps 45-48 of the main Procedure.

    PAUSE POINT: The RNA pellet in 75% ethanol can be stored up to a week at -80°C.

24. Re-dissolve the RNA pellet in 5 µl DEPC H_2O, and heat denature briefly in 65°C heat block for 20 s, then place on ice. Proceed from step 60 of the main Procedure.

Timing

Step 1, cell culture: 24 h

Step 2, sample preparation (nuclei isolation/cell permeabalization): 1 h

Steps 3-7, nuclear Run-On: 1.5 h

Steps 8-17, RNA extraction: 1 h

Steps 18-22, RNA fragmentation by base hydrolysis: 0.5 h

Steps 23-48, biotin RNA enrichment: 3 h

Steps 49-54, 3’ RNA adaptor ligation: 0.5 h hands-on, 4.5 h to overnight total (suggested end of the first day)

Steps 55 and 56, second biotin RNA enrichment: 3 h

Steps 57–63, enzymatic modification of the RNA 5′ ends: 3.5–4 h

Steps 64–74, hydroxyl repair (PRO-seq only): 2 h

Steps 75-80, 5’ RNA adaptor ligation: 4.5 h (0.5 h hands-on and 4 h incubation; can be stored overnight (suggested end of the second day)

Steps 81 and 82, third biotin RNA enrichment: 3 h

Steps 83-92, reverse transcription: 2 h

Steps 93-96, test PCR amplification: 2 h

Steps 97-101, test PCR gel analysis: 2 h (Suggested end of the third day)

Steps 102-110, full-scale PCR amplification: 2.5 h

Steps 111-146, library size selection by PAGE: 5 h – 1 day

Step 147, high-throughput sequencing, 24 h

Steps 148–150, data analysis: variable

Box 1, degradation of RNA containing 5′-monophosphate and removal of 5′-triphosphate and monophosphate from the RNA for PRO-cap: 4 h

Supplementary Tables

Supplementary Table 1. Nuclear Run-On reaction mix (step 4)

Reagents	**1× volume (µl)**	**Scaled volume (µl)**
NRO master mix	28	
Biotin NTP mix (+rNTP/H_2O)	20	
RNase inhibitor	2	
Sarkosyl (2%)	50	
Total	**100**	

Supplementary Table 2. RNA adaptor ligation mix (steps 52, 78)

Reagents	**1× volume (µl)**	**Scaled volume (µl)**
T4 RNA ligase buffer (10x)	1	
ATP (10 mM)	1	
50 % PEG	2	
RNase inhibitor	1	
T4 RNA ligase I	1	
Total	**6**	

Supplementary Table 3. 5'-phosphate-dependent exonuclease enzyme mix (Box 1, step 1, for PRO-cap only)

Reagents	**1× volume (µl)**	**Scaled volume (µl)**
DEPC H_2O	2.5	
Buffer A (10x)	1	
RNase inhibitor	0.5	
5'-phosphate-dependent exonuclease	1	
Total	**5**	

Supplementary Table 4. Alkaline phosphatase enzyme mix (Box 1, step 14, for PRO-cap only)

Reagents	**1× volume (µl)**	**Scaled volume (µl)**
DEPC H_2O	3	
Alkaline phosphatase buffer (10x)	1	
RNase inhibitor	0.5	
Alkaline phosphatase	0.5	
Total	**5**	

Supplementary Table 5. 5’ cap repair enzyme mix (step 60)

Reagents	**1× volume (µl)**	**Scaled volume (µl)**
DEPC H_2O	3/2.5	
TAP buffer/ThermoPol Reaction buffer (10x)	1	
RNase inhibitor	0.5	
TAP/RppH enzyme	0.5/1	
Total	**5**	

Supplementary Table 6. Polynucleotide kinase (PNK) enzyme mix (step 64, for PRO-seq only)

Reagents	**1× volume (µl)**	**Scaled volume (µl)**
DEPC H_2O	65	
10× PNK buffer	10	
ATP (10 mM)	10	
RNase inhibitor	2.5	
T4 PNK enzyme	2.5	
Total	**90**	

Supplementary Table 7. Reverse transcription buffer mix (step 87)

Reagents	1× volume (µl)	Scaled volume (µl)
First strand buffer (5x)	4	
DTT (0.1 M)	1	
RNase inhibitor	1	
Total	**6**	

Supplementary Table 8. Test PCR amplification mix (step 94)

Reagents	1× volume (µl)	Scaled volume (µl)
H_2O	5	
High Fidelity (HF) buffer (5x)	4	
Betaine (5 M)	4	
dNTP mix (12.5 mM each)	0.4	
RP1 primer (25 µM)	0.2	
RPI-1 primer (25 µM)	0.2	
Phusion DNA polymerase	0.	
Total	**14**	

Supplementary Table 9. Full-scale PCR amplification mix (step 102)

Reagents	1× volume (µl)	Scaled volume (µl)
H_2O	3	
High Fidelity (HF) buffer (5x)	10	
Betaine (5 M)	10	
dNTP mix (12.5 mM each)	1	
RP1 primer (25 µM)	0.5	
Phusion DNA polymerase	1	
Total	**25**	

REFERENCES

Adelman, K., and Lis, J.T. (2012). Promoter-proximal pausing of RNA polymerase II: emerging roles in metazoans. Nat. Rev. Genet. *13*, 720–731.

Andersson, R., Gebhard, C., Miguel-Escalada, I., Hoof, I., Bornholdt, J., Boyd, M., Chen, Y., Zhao, X., Schmidl, C., Suzuki, T., et al. (2014). An atlas of active enhancers across human cell types and tissues. Nature *507*, 455–461.

Carninci, P., Kvam, C., Kitamura, A., Ohsumi, T., Okazaki, Y., Itoh, M., Kamiya, M., Shibata, K., Sasaki, N., Izawa, M., et al. (1996). High-Efficiency Full-Length cDNA Cloning by Biotinylated CAP Trapper. Genomics *37*, 327–336.

Churchman, L.S., and Weissman, J.S. (2011). Nascent transcript sequencing visualizes transcription at nucleotide resolution. Nature *469*, 368–373.

Cock, P.J.A., Fields, C.J., Goto, N., Heuer, M.L., and Rice, P.M. (2010). The Sanger FASTQ file format for sequences with quality scores, and the Solexa/Illumina FASTQ variants.

Collart, M.A., and Oliviero, S. (2001). Preparation of Yeast RNA (Hoboken, NJ, USA: John Wiley & Sons, Inc.).

Core, L.J., Martins, A.L., Danko, C.G., Waters, C.T., Siepel, A., and Lis, J.T. (2014). Analysis of nascent RNA identifies a unified architecture of initiation regions at mammalian promoters and enhancers. Nat Genet *46*, 1311–1320.

Core, L.J., Waterfall, J.J., and Lis, J.T. (2008). Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. Science *322*, 1845–1848.

Core, L.J., Waterfall, J.J., Gilchrist, D.A., Fargo, D.C., Kwak, H., Adelman, K.,

and Lis, J.T. (2012). Defining the status of RNA polymerase at promoters. Cell Rep *2*, 1025–1035.
Fejes-Toth, K., Sotirova, V., Sachidanandam, R., Assaf, G., Hannon, G.J., Kapranov, P., Foissac, S., Willingham, A.T., Duttagupta, R., Dumais, E., et al. (2009). Post-transcriptional processing generates a diversity of 5′-modified long and short RNAs. Nature *457*, 1028–1032.

Forrest, A.R.R., Kawaji, H., Rehli, M., Kenneth Baillie, J., de Hoon, M.J.L., Haberle, V., Lassmann, T., Kulakovskiy, I.V., Lizio, M., Itoh, M., et al. (2014). A promoter-level mammalian expression atlas. Nature *507*, 462–470.

Fu, G.K., Xu, W., Wilhelmy, J., Mindrinos, M.N., Davis, R.W., Xiao, W., and Fodor, S.P.A. (2014). Molecular indexing enables quantitative targeted RNA sequencing and reveals poor efficiencies in standard library preparations. Proceedings of the National Academy of Sciences *111*, 1891–1896.

Fuda, N.J., Ardehali, M.B., and Lis, J.T. (2009). Defining mechanisms that regulate RNA polymerase II transcription in vivo. Nature *461*, 186–192.
García-Martínez, J., Aranda, A., and Pérez-Ortín, J.E. (2004). Genomic run-on evaluates transcription rates for all yeast genes and identifies gene regulatory mechanisms. Molecular Cell *15*, 303–313.

Hah, N., Murakami, S., Nagari, A., Danko, C.G., and Kraus, W.L. (2013). Enhancer transcripts mark active estrogen receptor binding sites. Genome Research *23*, 1210–1223.

Hah, N., Danko, C.G., Core, L., Waterfall, J.J., Siepel, A., Lis, J.T., and Kraus, W.L. (2011). A rapid, extensive, and transient transcriptional response to estrogen signaling in breast cancer cells. Cell *145*, 622–634.

Heinz, S., Romanoski, C.E., Benner, C., and Glass, C.K. (2015). The selection and function of cell type-specific enhancers. Nat Rev Mol Cell Biol *16*, 144–154.

Job, D., Job, D., Durand, R., Durand, R., Job, C., Job, C., Teissere, M., and

Teissere, M. (1984). Complex RNA chain elongation kinetics by wheat germ RNA polymerase II. Nucleic Acids Research *12*, 3303–3319.

Jonkers, I., Kwak, H., and Lis, J.T. (2014). Genome-wide dynamics of Pol II elongation and its interplay with promoter proximal pausing, chromatin, and exons. Elife *3*, e02407.

Kwak, H., and Lis, J.T. (2013). Control of Transcriptional Elongation. Annu. Rev. Genet. *47*, 483–508.

Kwak, H., Fuda, N.J., Core, L.J., and Lis, J.T. (2013). Precise maps of RNA polymerase reveal how promoters direct initiation and pausing. Science *339*, 950–953.

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol.

Larschan, E., Bishop, E.P., Kharchenko, P.V., Core, L.J., Lis, J.T., Park, P.J., and Kuroda, M.I. (2011). X chromosome dosage compensation via enhanced transcriptional elongation in Drosophila. Nature *471*, 115–118.

Larson, M.H., Mooney, R.A., Peters, J.M., Windgassen, T., Nayak, D., Gross, C.A., Block, S.M., Greenleaf, W.J., Landick, R., and Weissman, J.S. (2014). A pause sequence enriched at translation start sites drives transcription dynamics in vivo. Science *344*, 1042–1047.

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics *25*, 1754–1760.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078–2079.

Li, J., Liu, Y., Rhee, H.S., Ghosh, S.K.B., Bai, L., Pugh, B.F., and Gilmour, D.S. (2013). Kinetic Competition between Elongation Rate and Binding of NELF Controls Promoter-Proximal Pausing. Molecular Cell *50*, 711–722.

Mahat, D.B., Salamanca, H.H., Duarte, F.M., Danko, C.G., and Lis, J.T. (2016). Mammalian Heat Shock Response and Mechanisms Underlying Its Genome-wide Transcriptional Regulation. Molecular Cell *62*, 63–78.
Marcel, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet Journal *17*, 10–12.

Mayer, A., di Iulio, J., Maleri, S., Eser, U., Vierstra, J., Reynolds, A., Sandstrom, R., Stamatoyannopoulos, J.A., and Churchman, L.S. (2015). Native elongating transcript sequencing reveals human transcriptional activity at nucleotide resolution. Cell *161*, 541–554.

Min, I.M., Waterfall, J.J., Core, L.J., Munroe, R.J., Schimenti, J., and Lis, J.T. (2011). Regulating RNA polymerase pausing and transcription elongation in embryonic stem cells. Genes & Development *25*, 742–754.

Nojima, T., Gomes, T., Grosso, A.R.F., Kimura, H., Dye, M.J., Dhir, S., Carmo-Fonseca, M., and Proudfoot, N.J. (2015). Mammalian NET-Seq Reveals Genome-wide Nascent Transcription Coupled to RNA Processing. Cell *161*, 526–540.

Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics *26*, 841–842.

Rhee, H.S., and Pugh, B.F. (2012). Genome-wide structure and organization of eukaryotic pre-initiation complexes. Nature *483*, 295–301.
Seila, A.C., Calabrese, J.M., Levine, S.S., Yeo, G.W., Rahl, P.B., Flynn,

R.A., Young, R.A., and Sharp, P.A. (2008). Divergent transcription from active promoters. Science *322*, 1849–1851.

Vahedi, G., Kanno, Y., Furumoto, Y., Jiang, K., Parker, S.C.J., Erdos, M.R., Davis, S.R., Roychoudhuri, R., Restifo, N.P., Gadina, M., et al. (2015). Super-enhancers delineate disease-associated regulatory nodes in T cells. Nature *520*, 558–562.

Wang, D., Garcia-Bassets, I., Benner, C., Li, W., Su, X., Zhou, Y., Qiu, J., Liu, W., Kaikkonen, M.U., Ohgi, K.A., et al. (2011). Reprogramming transcription by distinct classes of enhancers functionally defined by eRNA. Nature *474*, 390–394.

Weber, C.M., Ramachandran, S., and Henikoff, S. (2014). Nucleosomes are context-specific, H2A.Z-modulated barriers to RNA polymerase. Molecular Cell *53*, 819–830.

APPENDIX B[e]

ENHANCED PRO-seq (ePRO-seq) FOR EFFICIENT AND HIGH-THROUGHPUT LIBRARY PREPARATION WITHOUT PCR AND SIZE-SELECTION BIASES

Development of ePRO-seq

PRO-seq maps the location of active RNA polymerases genome-wide at high resolution (Kwak et al., 2013). This assay has many attractive features over other genomic assays that also map RNA polymerases genome-wide. The recently published PRO-seq protocol (Mahat et al., 2016b) highlights advantages and limitations of this assay, and explains the procedure in great detail. However, PRO-seq can be laborious and takes 4 to 5 days. Library preparation of multiple samples is individually handled, increasing the likelihood of error. Once RNA adaptors on both ends of nascent RNA are ligated, PRO-seq library is amplified using PCR to generate quantifiable amount for optimal loading in the sequencing machine and size-selected using PAGE gel for optimal cluster generation and sequencing.

---

[e] Sections similar to PRO-seq are adapted from Mahat, D. B.*, Kwak, H.*, Booth, G. T., Jonkers, I. H., Danko, C. G., Patel, R. K., Waters, C.T., Munson, K., Core, L.J., & Lis, J.T. (2016). Base-pair-resolution genome-wide mapping of active RNA polymerases using precision nuclear run-on (PRO-seq). *Nature Protocols*, *11*(8), 1455–1476. *denotes equal contribution. Reprinted with permission from Nature Publishing.

The long library-preparation procedure and the likelihood of error due to extensive handling have been a source of intimidation for many young scientists who want to use this highly sensitive and robust genomic assay. The biases associated with PCR (Aird et al., 2011; Jones et al., 2015) has limited the ability of POR-seq assay to detect the transcription regulation in gene body manifested through pause. Similarly, size-selection bias (Poptsova et al., 2014) restrict the detection and interpretation of shorter promoter-proximal-paused reads. To address these limitations and simplify library preparation procedure, we developed an enhanced version of PRO-seq called ePRO-seq. This new method supports high-throughput library preparation by enabling sample mixing on the beginning of second day and reduces the library preparation time to 2.5 days. ePRO-seq library does not require PCR amplification and size selection, and thus eliminates the biases associated with these two steps common in genomic library preparation. By reducing time and minimizing handling, ePRO-seq saves reagents and reduces cost of library preparation making the enhanced version of PRO-seq a relatively easy, efficient, and rigorous genomic assay to map the transcriptionally engaged RNA polymerases genome-wide at base-pair resolution.

Overview of ePRO-seq

The major steps in ePRO-seq are similar to PRO-seq. Cell-culture, sample preparation, nuclear-run-on, RNA extraction and fragmentation by

base hydrolysis, and first biotin enrichment is identical. The process of 3' RNA adaptor ligation at the end of the first day is similar except for the difference in sequence. Unlike the 3' RNA adaptor VRA3 in PRO-seq, ePRO-seq uses PCRfreeRC3_BRn (n stands for barcode index), which contains in-line TRU-seq small RNA barcode with an additional G nucleotide at the very end of the 5' end. The presence of in-line barcode in 3' RNA adaptor allows for multiplexing of libraries on the second day and eliminates the necessity to carry each library in individual tube for multiple days. During library sequencing, the sequencing primer anneals upstream of the in-line barcode. Thus, the first six nucleotides of each read is the barcode sequence, which is used at post-processing step to parse sequences into individual libraries. The addition of a G nucleotide as the first base on the 5' end of 3' RNA adaptor in front of the barcode, which is also the nucleotide that gets directly ligated to the 3' end of nascent RNA, ensures uniformity in the ligating nucleotide as in PRO-seq and eliminates ligation bias that could be caused by different end-nucleotide of barcodes. This G is also removed at post-processing step prior to mapping of the reads to the genome.

The procedures on the second day of ePRO-seq are also similar to PRO-seq. One significant improvement in ePRO-seq is the ability to pool differentially barcoded libraries after the second biotin RNA enrichment step. For example, 48 ePRO-seq libraries made by using 12 barcodes can be pooled into 4 tubes (12 differentially barcodes libraries in each tube). The

steps of de-capping and hydroxyl-repair are identical, although the reaction volume may be scaled up to match the number of libraries pooled. The 5' RNA adaptor used in ePRO-seq (PCRfreeRC5 for single-end sequencing and 5Adapt_Paired_ePROseq for paired-end sequencing) is also different from VRA5 used in PRO-seq, however, the reaction is carried out similarly.

On the third day, the third biotin RNA enrichment and reverse transcription is carried out just like in PRO-seq. However, degradation of RNA after reverse transcription is critical in ePRO-seq. The resulting cDNA is ready for sequencing and is submitted to sequencing facility without further processing. All subsequent steps in PRO-seq such as test PCR amplification, gel analysis of test PCR, full-scale PCR amplification, and library-size selection by PAGE are eliminated in ePRO-seq. The removal of these steps significantly reduces the library preparation time (from 4-5 days to 2.5 days), saves reagents, and cuts cost. More importantly, it eliminates the biases associated with PCR amplification of genomic library and size selection for optimal sequencing.

A schematic of steps involved in ePRO-seq and their comparison with steps involved in PRO-seq is shown in Figure B.1. Major differences remain in RNA adaptors sequence and the placement of barcode.

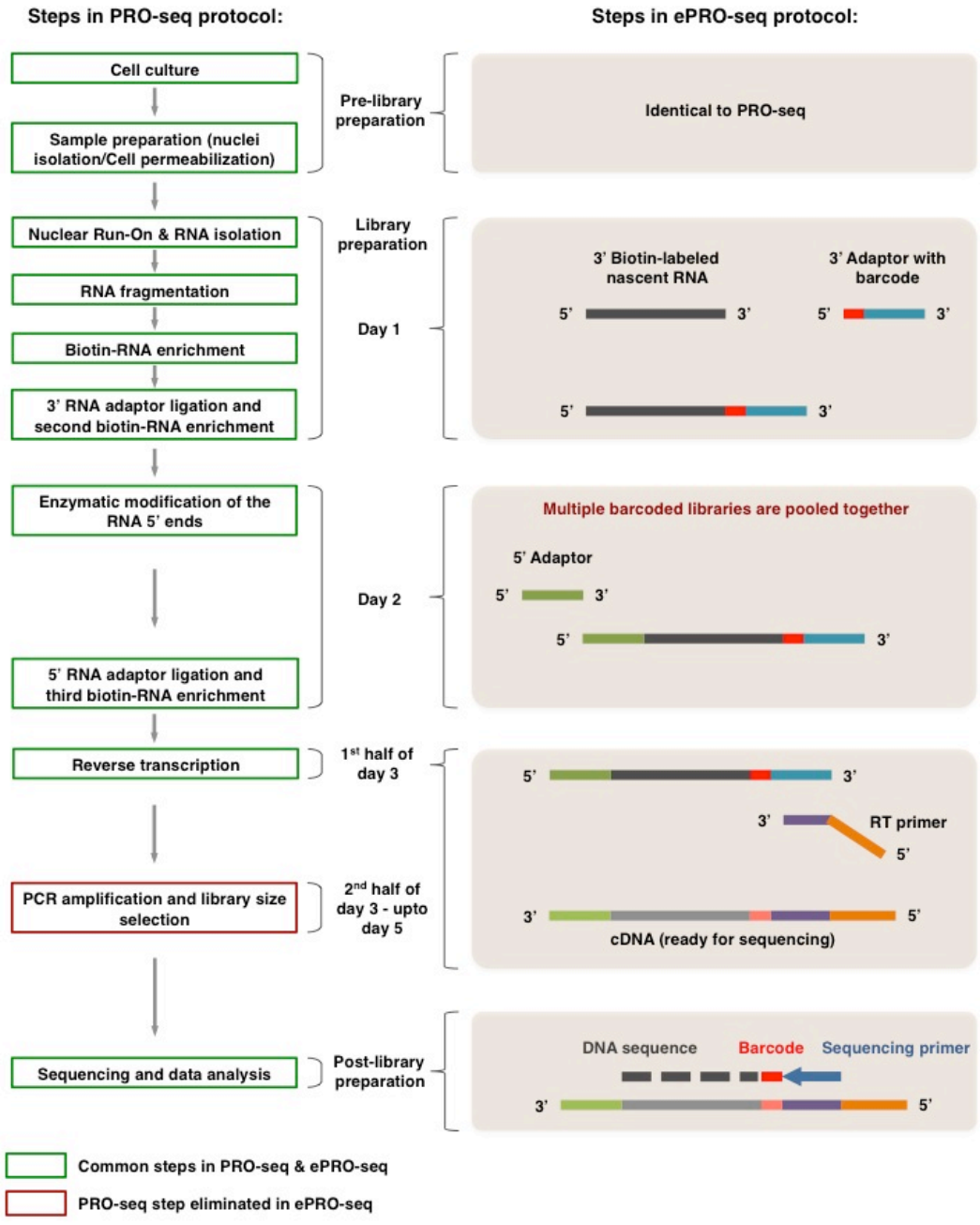


Figure B.1. Overview of ePRO-seq and its comparison to PRO-seq
Steps in PRO-seq that correspond to steps in ePRO-seq are shown side-by-side. PRO-seq steps in green boxes are present in ePRO-seq and PRO-seq steps in red box are eliminated from ePRO-seq. Compared to 4-5 days required for PRO-seq library preparation, ePRO-seq library can be prepared in 2.5 days. The RNA adaptors sequences are different between PRO-seq and ePRO-seq and the multiplesing barcode in ePRO-seq is present as in-line barcode in 3' RNA adaptor.

Experimental Design

Cells

We have successfully made ePRO-seq libraries with MEFs. However, PRO-Seq libraries are routinely made from plant (unpublished, G.T.B.), yeast (unpublished, G.T.B.), *Drosophila* (Kwak et al., 2013), and mammalian cells (Core et al., 2014; Mahat et al., 2016a), and ePRO-seq should be easily adaptable to these cells. In general, the higher the number of cells the better the ePRO-seq read coverage of the genome. Although a minimum of 5-10 million nuclei or permeabilized cells is recommended for a single ePRO-seq library regardless of cell type, the option of pooling of multiple differentially barcoded libraries after second bead binding may tolerate lower number of cells in individual library. The number of barcodes available is the only factor that limits the extent of library pooling. In principle, the application of PRO-seq and ePRO-seq in yeast, including *S. pombe* and *S. cerevisiae*, is very similar to that of other organisms; however, some alterations are required in yeast cell permeabilization (García-Martínez et al., 2004), run-on reaction, and post run-on RNA extraction procedures (Collart and Oliviero, 2001); required modifications for yeast are indicated in the appropriate steps of the Procedure.

Sample preparation

Isolation of nuclei for nuclear run-on is a critical step in the procedure not only to preserve the enzymatic activity of the RNA polymerase, but also to capture the precise position of the RNA polymerase on genes. Starting with

10-20 million cells per library is recommended considering the efficiency of nuclei isolation process (~50%). The whole process should take place in the cold room on ice as far as is possible. Isolated nuclei can be resuspended in the glycerol-containing storage buffer, and quickly frozen in liquid nitrogen for long term storage at -80°C. We have used permeabilized cells in PRO-seq and ePRO-seq as an alternative to isolating nuclei, making handling easier and reducing loss of sample; cell permeabilization has a much higher efficiency (~90%) than nuclei isolation. Cell permeabilization conditions may differ between cell types and may need to be optimized; we provide a general method for permeabilization in the Procedure as well as a version optimized for yeast cells.

Spike-in for library normalization

Disproportionate loss of RNA and/or cDNA can occur during multiple stages of the PRO-seq library preparation, which spans 4-5 days and involves several handling steps. Even with the use of identical starting material, uneven loss of libraries could affect the genome-wide RNA polymerase density between libraries. To control for handling effects on library yield, a small fraction (1-5%) of cells with a distinct genome can be added during library preparation; adding an identical number of spike-in cells to different libraries enables normalization between different conditions. We have used *Schizosaccharomyces pombe* to normalize *Saccharomyces cerevisiae* and vice versa, and *Drosophila* cells to normalize mammalian cells and vice versa.

When using the cell-permeabilization approach, spike-in cells should be added and permeabilized together with the experimental cells. For the nuclei isolation approach, spike-in cells should be added to the experimental cells prior to nuclei isolation and dounced together.

Nuclear run-on

In PRO-seq, biotin-NTPs are used as the nuclear run-on substrates. The K_m of each of the NTPs as substrates for RNA polymerase lie in the range of 1–20 μM (Job et al., 1984). Therefore, final substrate concentration greater than 1-20 μM range (~25 μM) is, in general, sufficient for each biotin-NTP substrate. Depending on the purpose of the experiment, biotin-NTP substrates can be supplied in different combinations: individual biotin run-on, 4 biotin run-on, 2 biotin run-on or 1 biotin run-on.

- Individual-biotin run-on: To obtain the most precise mapping of the RNA polymerase, four separate PRO-seq libraries are made, each supplied with only one type of biotin-NTP in the nuclear run-on reaction. This ensures that the RNA polymerase adds only one, or at most a few (when the polymerase is positioned at multiple stretches of same nucleotide) biotin-NTPs to the nascent RNA. In this case, 4 times more sample is required.
- 4-Biotin run-on: we found that all 4 biotin-NTPs can be supplied in a single reaction and the Pol II only incorporates one or at most a few bases, giving an equivalent resolution to Individual-biotin run-on. The reason for this is

unclear, but we speculate that steric hindrance in the active site of RNA polymerase prevents incorporation of multiple biotinylated nucleotides.

- 2-Biotin run-on: When the amount of sample or the cost is limiting, unlabeled NTPs can be used in combination with biotin NTPs. Use of biotinylated purine nucleotides (biotin-ATP, biotin-GTP) is more costly than that of the pyrimidine nucleotides (biotin-CTP, biotin-UTP). A combination of biotin-CTP, biotin-UTP, ATP, and GTP can be used in run-on reaction, providing reasonable resolution and cost.
- 1-Biotin run-on: If a longer run-on RNA is preferred, combinations of biotin-CTP with unlabeled CTP, UTP, ATP, and GTP can be used effectively in a biotin-NTP form of GRO-seq. This approach can be useful for increasing sequencing coverage of RNA polymerases that reside near the TSSs. While most transcriptionally-engaged RNA polymerases near the 5' ends reside between 30-60 nucleotides from the TSS (Kwak et al., 2013), RNA polymerases closer to the TSS may fail to map uniquely. Additionally, the longer run-on of nascent RNAs may be desired for distinguishing allele-specific nascent transcription.

Multiplexing

The presence of multiplexing barcodes in the 3' RNA adaptor enables pooling of libraries after the second bead binding (first step on Day 2). This significantly reduces the number of tubes and simplifies handling. Only the number of barcodes being used dictates the extent of pooling.

Materials

Reagents

CRITICAL: Extreme care should be taken to avoid nuclease contamination. Use nuclease-free reagents and change gloves routinely.

- Appropriate cell line(s) e.g. K562, GM12878, MCF7, Hela, embryonic stem cells, mouse embryonic fibroblasts, mouse 3T3 cells, *Drosophila* S2, yeast.

  CAUTION: Before use, cells should be checked for contamination.

Chemical stocks

- Diethyl pyrocarbonate (DEPC; Sigma-Aldrich, cat. no. D5758)

  CAUTION: DEPC is toxic and harmful. Proper eyeshield, faceshield, full-face respirator, and gloves are required while handling DEPC.
- Sodium chloride, NaCl (Sigma-Aldrich, cat. no. S9888)
- Potassium chloride, KCl (Avantor, cat. no. 6858-04)
- Magnesium chloride, $MgCl_2$ (Avantor, cat. no. 5958-04)
- Sucrose (Sigma-Aldrich, cat. no. S0389)
- Calcium chloride, $CaCl_2$ (Sigma-Aldrich, cat. no. C1016)
- Magnesium acetate, $MgAc_2$ (Sigma-Aldrich, cat. no. M5661)
- Ammonium acetate, NH_4Ac (Sigma-Aldrich, cat. no. A1542)
- Sodium acetate NaOAc (Sigma-Aldrich, cat. no. S2889)
- EDTA (Sigma-Aldrich, cat. no. E9884)
- EGTA (Sigma-Aldrich, cat. no. E3889)
- Protease inhibitor cocktail, EDTA-free (Roche, cat. no. 11873580001)

- Sodium hydroxide, NaOH (Avantor, cat. no. 7708-10)
- Triton X-100, (Calbiochem, cat. no. 9410)
- Nonidet P40 (NP40) Substitute, (Sigma-Adlrich, cat. no. 11332473001)
- Sarkosyl (Sigma-Aldrich, cat. no. L5125)
- Tween-20 (Sigma-Aldrich, cat. no. P9416)
- Phosphate buffer saline, PBS pH 7.4 (Gibco, cat. no. 10010031).
- TRIS (Avantor, cat. no. 4109-02)
- Hydrochloric acid, HCl (Avantor, cat. no. 4613-05)
- DTT (Sigma-Aldrich, cat. no. D0632)
- Glycerol (Sigma-Aldrich, cat. no. G5516)

Biotin Nuclear Run-On and enrichment

- Biotin-11-ATP (PerkinElmer, cat. no. NEL544001EA)
- Biotin-11-CTP (PerkinElmer, cat. no. NEL542001EA)
- Biotin-11-GTP (PerkinElmer, cat. no. NEL545001EA)
- Biotin-11-UTP (PerkinElmer, cat. no. NEL543001EA)
- ATP, 10mM (Roche, cat. no. 11 277 057 001)
- GTP, 10mM (Roche, cat. no. 11 277 057 001)
- UTP, 10mM (Roche, cat. no. 11 277 057 001)
- P-30 column, RNase free (BIORAD, cat. no. 732-6250)
- Streptavidin M280 beads (Invitrogen, cat. no. 112.06D)

Reagents for nucleic acid extraction

- Trizol (Ambion, cat. no. 115596018)

CAUTION: Trizol is harmful and contact with skin, eye or inhalation should be avoided. Use it inside a fume hood.

- Trizol LS (Ambion, cat. no. 10296028)

  CAUTION: Trizol is harmful and contact with skin, eye or inhalation should be avoided. Use it inside a fume hood.
- Chloroform (Calbiochem, cat. no. 3150)
- GlycoBlue (Ambion, cat. no. AM9515)
- Ethanol, 100% (PHARMCO-AAPER, cat. no. 111000200)
- Ethanol, 75%(vol/vol)
- Phenol:Chloroform, Tris buffered (Thermo Scientific, cat. no. 17909)

  CAUTION: Phenol:Chloroform is harmful and contact with skin, eye or inhalation should be avoided. Use it inside a fume hood.
- Phenol, (Ambion, cat. no. 9700).

  CAUTION: Phenol is harmful and contact with skin, eye or inhalation should be avoided. Use it inside a fume hood.

Enzymes and recombinant protein reagents

- RNase inhibitor, 40 units/μl (Ambion, cat. no. AM2696)
- T4 RNA ligase I, 10 units/μl (NEB, cat. no. M0204). Supplied with 10× T4 RNA ligase buffer, 10 mM ATP, and PEG, 50%(wt/vol).
- 5´-phosphate-dependent exonuclease, 1 unit/μl (Epicenter, cat. no. TER51020) (required for PRO-cap only). Supplied with 10× reaction buffer A.

- Alkaline phosphatase, 10 units/μl (NEB, cat. no. M0290) (required for PRO-cap only). Supplied with 10× Alkaline phosphatase buffer. Alternatively, Antarctic phosphatase, 5 units/μl (NEB, cat. no. M0289) can be used.
- Tobacco acid pyrophosphatase, 10 units/μl (TAP) (Epicenter, cat. no. T19500). Supplied with 10× TAP buffer. Alternatively, RNA 5' Pyrophosphohydrolase, 5 units/μl (RppH) (NEB, cat. no. M0356S) can be used with ThermoPol Reaction buffer (NEB, cat. no. B9004S).
- T4 polynucleotide kinase, 10 units/μl (PNK) (NEB, cat. no. M0201) (required for PRO-seq only). Supplied with10× PNK buffer.
- Superscript III reverse transcriptase (Invitrogen, cat. no. 56575). Supplied with 5× first strand buffer, and 0.1M DTT.
- dNTP mix, 12.5 mM each (Roche, cat. no. 03 622 614 001)
- RNase cocktail, 0.5 U/μl of RNase A and 20 U/μl of RNase T1 (Ambion, AM2286)
- RNase H, 5 U per μl (NEB, cat. no. M0297S)
- RNA and DNA oligos. (Custom synthesis from IDT DNA, RNase-free HPLC purified) See Table 1 and Reagent Setup for details. Further information about barcoding and sequencing indexes can be found at http://support.illumina.com/content/dam/illumina-support/documents/documentation/chemistry_documentation/samplepreps_truseq/truseqsampleprep/truseq-library-prep-pooling-guide-15042173-01.pdf

Equipment

- 2 heat blocks, one set at 37°C and the other at 65°C, each filled with water equilibrated at the appropriate temperature.
- Dounce homogenizer (Wheaton scientific, cat. no. 357546) (for nuclei isolation)
- Magnetic stand for Streptavidin magnetic beads (Invitrogen, cat. no. K1585-01)
- Rotating stand (Thermo Barnstead Labquake rotator, cat. no. 415110)
- Refrigerated centrifuge (Eppendorf, cat. no. 5417R)
- Microcentrifuge (Eppendorf, cat. no. 5415D)

Reagent Setup

CRITICAL: All reagents, solutions, and buffers should be made with DEPC-treated water

- **DEPC-H_2O** Add 0.1%(vol/vol) DEPC to H_2O. Mix overnight then autoclave and filter-sterilize the solution with a 0.22 μm filter. DEPC-H_2O can be prepared in advance and stored at room temperature (25^0C) for up to a year.

  CAUTION: DEPC is toxic and harmful. Proper eyeshield, faceshield, full-face respirator, and gloves are required while handling DEPC.
- **5M NaCl** Dissolve 14.61 g NaCl in 50 ml H_2O with 0.1%(vol/vol) DEPC. Mix overnight and then autoclave and filter.

CRITICAL: 5M NaCl can be stored at room temperature (25^0C) for up to a year.

- **4M KCl** Dissolve 3.73 g KCl in 50 ml H_2O with 0.1%(vol/vol) DEPC, mix overnight then autoclave and filter-sterilize.

  CRITICAL: 4M KCl can be stored at room temperature for up to a year.

- **1M $MgCl_2$** Dissolve 4.76 g $MgCl_2$ in 50 ml H_2O with 0.1%(vol/vol) DEPC, mix overnight then autoclave and filter-sterilize.

  CRITICAL: 1M $MgCl_2$ can be stored at room temperature for up to a year.

- **1M Sucrose** Dissolve 171.15 g Sucrose in 500 ml H_2O with 0.1%(vol/vol) DEPC, mix overnight then autoclave and filter-sterilize.

  CRITICAL: 1M Sucrose can be stored at room temperature for up to a year.

- **1M $CaCl_2$** Dissolve 5.55 g $CaCl_2$ in 50 ml H_2O with 0.1%(vol/vol) DEPC, mix overnight then autoclave and filter-sterilize.

  CRITICAL: 1M $CaCl_2$ can be stored at room temperature for up to a year.

- **1M $MgAc_2$** Dissolve 7.12 g $MgAc_2$ in 50 ml H_2O with 0.1%(vol/vol) DEPC, mix overnight then autoclave and filter-sterilize.

  CRITICAL: 1M $MgAc_2$ can be stored at room temperature for up to a year.

- **1M NH_4Ac** Dissolve 3.85 g NH_4Ac in 50 ml H_2O with 0.1%(vol/vol) DEPC, mix overnight then autoclave and filter-sterilize.

  CRITICAL: 1M NH_4Ac can be stored at room temperature for up to a year.

- **1M NaOAc, pH 5.3** Dissolve 4.1 g NaOAc in 50 ml H_2O and pH to 5.3, add 0.1%(vol/vol) DEPC, mix overnight then autoclave and filter-sterilize.

  CRITICAL: 1M NaOAc can be stored at room temperature for up to a year.
- **0.5M EDTA** Dissolve 29.22 g EDTA in 100 ml DEPC treated H_2O, then autoclave and filter-sterilize.

  CRITICAL: 0.5M EDTA can be stored at room temperature for up to a year.
- **0.1M EGTA** Dissolve 19.02 g EGTA in 50 ml DEPC treated H_2O, then autoclave and filter-sterilize.

  CRITICAL: 0.1M EGTA can be stored at room temperature for up to a year.
- **1N NaOH** Dissolve 2 g NaOH in 50 ml DEPC treated H_2O, filter-sterilize.

  CRITICAL: 1N NaOH can be prepared in advance in 50 ul aliquots, stored at -80°C for up to a year. Use freshly thawed aliquot each time.

  CAUTION: NaOH is corrosive and contact with skin, eye or inhalation should be avoided.
- **10% Triton X-100** Dissolve 5 ml of Triton X-100 in 45 ml DEPC H_2O and filter-sterilize.

  CRITICAL: 10% Triton X-100 can be stored at room temperature for up to a year.
- **10% NP40** Dissolve 5 ml of NP40 in 45 ml DEPC H_2O and filter-sterilize.

  CRITICAL: 10% NP40 can be stored at room temperature for up to a year.
- **2% Sarkosyl** Dissolve 1 g of Sarkosyl in 50 ml DEPC H_2O and filter-sterilize. CRITICAL: 2% Sarkosyl can be stored at room temperature for up

to a year. CAUTION: Sarkosyl is an irritant and contact with skin, eye or inhalation should be avoided.

- **1% Tween-20** Dissolve 1 ml of Tween-20 in 49 ml DEPC H_2O and filter-sterilize. CRITICAL: 1% Tween-20 can be stored at room temperature for up to a year.
- **1M Tris-HCl, pH 6.8** Dissolve 6.06 g TRIS base in 50 ml DEPC H_2O, pH to 6.8 with HCl then autoclave and filter-sterilize.

  CRITICAL: The buffer can be stored at room temperature for up to a year.
- **1M Tris-HCl, pH 7.4** Dissolve 6.06 g TRIS base in 50 ml DEPC H_2O, pH to 7.4 with HCl then autoclave and filter-sterilize.

  CRITICAL: The buffer can be stored at room temperature for up to a year.
- **1M Tris-HCl, pH 8.0** Dissolve 6.06 g TRIS base in 50 ml DEPC H_2O, pH to 8.0 with HCl then autoclave and filter-sterilize.

  CRITICAL: The buffer can be stored at room temperature for up to a year.
- **1M DTT** Dissolve 1.54 g DTT in 10 ml DEPC H_2O and filter-sterilize.

  CRITICAL: 1M DTT can be prepared and stored at -20^{0}C for up to a year.
- **1mM Biotin-11-CTP** Mix 10 μl of 10 mM stock in 90 μl DEPC H_2O to get 1 mM dilution.

  CRITICAL: The buffer can be stored at 4^{0}C for up to a year.
- **1mM Biotin-11-UTP** Mix 10 μl of 10 mM stock in 90 μl DEPC H_2O to get 1 mM dilution.

  CRITICAL: The buffer can be stored at 4^{0}C for up to a year.

- **Douncing buffer (for nuclei isolation)** 10 mM Tris-HCl pH 7.4, 300mM sucrose, 3 mM $CaCl_2$, 2 mM $MgCl_2$, 0.1% Triton X-100, 0.5 mM DTT, 1 tablet of protease inhibitors cocktail per 50ml, 4 u/ml RNase inhibitor.

  CRITICAL: The buffer without DTT, protease inhibitors, and RNase inhibitor can be prepared and stored at 4^0C for up to a month. Add DTT, protease inhibitors, and RNase inhibitor immediately before use.
- **Permeabilization buffer (for non-yeast cells)** 10mM Tris-HCl pH 7.4, 300mM sucrose, 10mM KCl, 5mM $MgCl_2$, 1mM EGTA, 0.05% Tween-20, 0.1% Nonidet P40 substitute, 0.5 mM DTT, 1 tablet of protease inhibitors cocktail per 50ml, 4 u/ml RNase inhibitor.

  CRITICAL: The buffer without DTT, protease inhibitors, and RNase inhibitor can be prepared and stored at 4^0C for up to a month. Add DTT, protease inhibitors, and RNase inhibitor immediately before use.
- **Permeabilization buffer (for yeast cells)** 0.5% Sarkosyl, 0.5 mM DTT, 1 tablet of protease inhibitors cocktail per 50ml, 4 u/ml RNase inhibitor.

  CRITICAL: The buffer without DTT, protease inhibitors, and RNase inhibitor can be prepared and stored at 4^0C for up to a month. Add DTT, protease inhibitors, and RNase inhibitor immediately before use.
- **Storage buffer** 10 mM Tris-HCl pH 8.0, 25%(vol/vol) glycerol, 5 mM $MgCl_2$, 0.1 mM EDTA, 5 mM DTT.

  CRITICAL: The buffer without DTT can be prepared and stored at 4^0C for up to a month. Add fresh DTT immediately before use.

- **2x Nuclear run-on master mix (for non-yeast cells)** 10 mM Tris-HCl pH 8.0, 5 mM $MgCl_2$, 300 mM KCl, and 1 mM DTT.

  CRITICAL: The buffer without DTT can be prepared and stored at 4^0C for up to a month. Add fresh DTT immediately before use.
- **2x NRO master mix (for yeast cells)** 40 mM Tris-HCl, pH 7.7, 400 mM KCl, 64 mM $MgCl_2$, 1 mM DTT.)

  CRITICAL: The buffer without DTT can be prepared and stored at 4^0C for up to a month. Add fresh DTT immediately before use.
- **AES buffer (for yeast cells only)** 50 mM NaOAc pH 5.3, 10 mM EDTA, 1% SDS.
- **High-salt wash buffer** 50 mM Tris-HCl pH 7.4, 2 M NaCl, 0.5%(vol/vol) Triton X-100 in DEPC H_2O.

  CRITICAL: The buffer can be stored at 4^0C for up to a month.
- **Binding buffer** 10 mM Tris-HCl pH 7.4, 300 mM NaCl, 0.1%(vol/vol) Triton X-100 in DEPC H_2O.

  CRITICAL: The buffer can be stored at 4^0C for up to a month.
- **Low-salt wash buffer** 5 mM Tris-HCl pH 7.4, 0.1%(vol/vol) Triton X-100 in DEPC H_2O.

  CRITICAL: The buffer can be stored at 4^0C for up to a month.
- **Pre-washed streptavidin-coated magnetic beads** Take 90 µl of Streptavidin M280 beads per library. Place the beads on the magnetic separator for 1 min and discard the supernatant. Pre-wash by

resuspending in 0.1 N NaOH + 50 mM NaCl in DEPC H_2O for 1 min, place on the magnetic separator for 1 min, remove supernatant. Wash beads twice with 100 mM NaCl in DEPC H_2O. After removing the wash buffer, resuspend the beads in 150 µl of the Binding Buffer and make 3 aliquots of 50 µl each. Scale up accordingly when processing multiple samples. CRITICAL: The washed beads can be stored at 4^0C for up to a week.

- **DNA and RNA oligos** Oligos for PRO-seq and PRO-cap (Table B.1) should be dissolved in DEPC H_2O at a concentration of 100 mM. PCR primers should be dissolved in DEPC H_2O at a concentration of 25 mM. CRITICAL: The DNA and RNA oligos can be stored at -80^0C for years.

Software for data analysis

- Barcode splitter, sequence trimmer, adaptor clipper, and reverse complement software as a part of 'fastx' toolkit (http://hannonlab.cshl.edu/fastx_toolkit/commandline.html)
- Mapping or alignment software, such as 'bwa'(Li and Durbin, 2009) (https://sourceforge.net/projects/bio-bwa/files/) or 'bowtie'(Langmead et al., 2009) (https://sourceforge.net/projects/bowtie-bio/files/)
- Tools to generate coverage information, such as SAMtools(Li et al., 2009) (https://sourceforge.net/projects/samtools/files/) and BEDTools(Quinlan and Hall, 2010) (https://sourceforge.net/projects/bedtools/files/)

Table B.1. Oligonucleotides required for ePRO-seq.

Oligo name	Sequence (5' to 3')	Purpose	Comments
PCRfreeRC3_BRn	GNNNNNN GAUCGUC GGACUGU AGAACUCU GAAC-/Inverted dT/	RNA adaptor for ligation to the 3' end of nascent RNA	Reverse complement of sequencing primer for TRUseq Small RNA (26bp), with TRUseq indexes on 5' end plus a terminal G. The six Ns is in-line barcode (reverse complement of Illumina TRU-seq index). The 5' end of the adaptor is phosphorylated and the 3' end is protected by an inverted dT.
PCRfreeRC5	CAAGCAGA AGACGGC AUACGAGA U	RNA adaptor for ligation to the 5' end of nascent RNA for single-end sequencing.	RNA version of one of the primers in flow-cell (with the G'). The 5' end of the adaptor is not phosphorylated.
5Adapt_Paired_ePRO	CAAGCAGA AGACGGC AUACGAGA UGUCUCG UGGGCUC GGAGAUG UGUAUAAG AGACAG	RNA adaptor for ligation to the 5' end of nascent RNA for paired-end sequencing. Designed by Jacob M. Tome.	The first 24 bp is the RNA version of one of the primers in flow-cell (with the G') and the remaining 34 bp is the RNA version of the Read 2 sequencing primer in paired-end sequencing. The 5' end of the adaptor is not phosphorylated.
RP1	AATGATAC GGCGACC ACCGAGAT CTACACGT TCAGAGTT CTACAGTC CGA	DNA oligo for reverse transcription of adaptors-ligated nascent RNA	Same RP1 primer used in PRO-seq.
TestAmp3	CAAGCAGA AGACGGC AUACGAGA U	DNA oligo for amplification of both single-end and paired-end libraries (test amplification or emergency amplification)	One of the primers in flow-cell (with the G') with exact sequence.

Procedure

Cell culture l TIMING 24 h

1. Seed cells at a concentration that will enable them to reach ~80% confluency in 24 hours. For a PRO-seq experiment using adherent fibroblasts, 4-6 150mm cell culture dishes yield sufficient cells (~10^7 cells, see Experimental Design for further details). For yeast cells, plate them to ensure they are in the exponential phase of growth (OD_{600} = 0.5) at the time of harvest.

   CAUTION: Check cell lines for mycoplasma contamination before setting up the experiment.

Sample preparation l TIMING: 1 h

CRITICAL: Samples should be prepared in 4^0C to avoid unsolicited run-on.

2. Prepare samples by isolating nuclei (Option A) or by cell permeabilization (use Option C for yeast cells and Option B for other cell types). All centrifugation steps for sample preparation are performed in a cold centrifuge (4^0C) at 1000 g (unless stated otherwise) for 5 min.

   A. Nuclei isolation

   a. Harvest adherent cells by scraping and centrifuging, and non-adherent cells by centrifuging.
   b. Resuspend the cell pellet in 10 ml ice-cold PBS and centrifuge.
   c. Resuspend the cell pellet in ice-cold douncing buffer (1x10^6 cells/ml).

d. CRITICAL STEP: If using spike-in cells, add them at this point to the cells resuspended in douncing buffer.

e. Incubate for 5 min on ice and dounce 25 times using a dounce homogenizer.

f. Transfer the dounced nuclei to either a 15 or 50 ml conical tube and centrifuge the nuclei.

g. Wash twice by resuspending the pellet in 5 ml douncing buffer and centrifuging.

h. Resuspend the pellet in storage buffer (5-10x10^6 nuclei per 100 μl of storage buffer), flash freeze in liquid nitrogen, and store at -80^0C.

PAUSE POINT: The nuclei in storage buffer can be stored at -80^0C for up to 5 years.

B. Cell permeabilization (non-yeast cells)

a. Harvest adherent cells by scraping and centrifuging, and non-adherent cells by centrifuging.

b. Resuspend the cell pellet in 10 ml ice-cold PBS and centrifuge.

c. Resuspend the cell pellet in ice-cold permeabilization buffer (1x10^6 cells/ml).

CRITICAL STEP: Spike-in cells, if used, should be added to the cells resuspended in permeabilization buffer.

d. Incubate for 5 min on ice and centrifuge the permeabilized cells.

e. Wash twice by resuspending in 5 ml permeabilization buffer and centrifuging.

f. Resuspend the cell pellet in storage buffer (5-10x10^6 permeabilized cells per 100 μl of storage buffer), flash freeze in liquid nitrogen, and store in -80^0C.

PAUSE POINT: The permeabilized cells in storage buffer can be stored at -80^0C for up to 5 years.

C. Cell permeabilization (optimized for yeast)

a. Harvest exponentially growing yeast cells by centrifugation at 400 g.

b. Resuspend the cell pellet in 10 ml ice-cold PBS and centrifuge.

c. Resuspend the cell pellet in ice-cold yeast permeabilization buffer (1x10^6 cells/ml).

CRITICAL STEP: Spike-in cells, if used, should be added to the cells resuspended in yeast permeabilization buffer.

d. Incubate for 20 min on ice and centrifuge the cells at 400 g.

e. Resuspend the cell pellet in storage buffer (25-50x10^6 permeabilized cells per 100 μl of storage buffer), flash freeze in liquid nitrogen, and store in -80^0C.

PAUSE POINT: The permeabilized cells in storage buffer can be stored at -80^0C for up to 5 years.

Nuclear run-on l TIMING: 1.5 h

3. Prepare a 2x nuclear run-on (NRO) master mix; for non-yeast cells, prepare the master mix according to the first table, for yeast cells, use the second table.

Reagents	**Volume per reaction (μl)**	**Final concentration – 1x (in 200-μl reaction) (mM)**
Tris-Cl pH 8.0 (1 M)	1	5
$MgCl_2$ (1 M)	0.5	2.5
DTT (0.1 M)	1	0.5
KCl (4 M)	7.5	150
DEPC-H_2O	18	

Reagents	**Volume per reaction (μl)**	**Final concentration – 1x (in 200-μl reaction) (mM)**
Tris-Cl pH 7.7 (1 M)	4	20
$MgCl_2$ (1 M)	6.4	32
DTT (0.1 M)	1	0.5
KCl (4 M)	10	200
DEPC-H_2O	6.6	

4. Depending on the type of run-on experiment (see Experimental Design section), prepare a 2x reaction mix according to Option A (single biotin run-on), Option B (4 biotin run-on), Option C (2 biotin run-on) or Option D (1 biotin run-on). If processing multiple libraries at once, scale up accordingly.

A. Individual-biotin run-on 2x reaction mix

a. Transfer a 28 μl aliquot of NRO master mix to each of 4 separate microcentrifuge tubes.

b. Add 5 μl of biotin-11-ATP (1 mM) to one of the tubes containing NRO master mix. Label this mix “A”

c. Repeat step b for the remaining 3 biotin-11-NTPs (1 mM each) and the 3 remaining tubes containing NRO master mix and label them “C”, “G”, “U” accordingly.

d. Add 15 μl DEPC H_2O to all 4 tubes.

e. Add 2 μl of RNase inhibitor and 50 μl of 2% Sarkosyl to all 4 tubes.

From step e, each tube will be processed as a separate sample.

B. 4-Biotin run-on 2x reaction mix

a. Transfer a 28 μl aliquot of NRO master mix to a microcentrifuge tube.

b. Add 5 μl each of all 4 biotin-11-NTPs (1 mM each) to the NRO master mix aliquot.

c. Add 2 μl of RNase inhibitor and 50 μl of 2% Sarkosyl.

C. 2-Biotin run-on 2x reaction mix

a. Transfer a 28 μl aliquot of NRO master mix to a microcentrifuge tube.

b. Add 5 μl each of biotin-11-CTP (1 mM) and biotin-11-UTP (1 mM) to the NRO master mix aliquot.

c. Add 2.5 μl each of ATP (10 mM) and GTP (10 mM) to the mix.

d. Add 5 μl DEPC H_2O.

e. Add 2 μl of RNase inhibitor and 50 μl of 2% Sarkosyl.

D. 1-Biotin run-on 2x reaction mix

a. Transfer a 28 μl aliquot of NRO master mix to a microcentrifuge tube.

b. Add 5 μl of biotin-11-CTP (1 mM) to the NRO master mix aliquot.

c. Add 1 μl of CTP (0.05 mM) to the mix.

d. Add 2.5 μl each of ATP (10 mM), GTP (10 mM), and UTP (10 mM) to the mix.

e. Add 6.5 μl DEPC H_2O.

f. Add 2 μl of RNase inhibitor and pipette up and down several times.

g. Add 50 μl of 2% Sarkosyl and pipette up and down 15 times.

5. Preheat 100 μl of the appropriate 2x reaction mix prepared in step 4 to 37°C for mammalian cells or 30°C for yeast and insect cells.

6. Using a cut-off P200 pipette tip, add 100 μl nuclei or permeabilized cells (in storage buffer from step 2) to 100 μl of preheated 2x reaction mix, gently but thoroughly pipette the reaction 15 times, and place in a heat block at the appropriate temperature.

   CRITICAL STEP: Sarkosyl in the 2x reaction mix causes the run-on reaction to become very viscous (except for yeast). When adding the nuclei or permeabilized cells to the reaction mix and when mixing by pipetting up and down, use a wide bore pipette tip or cut the last centimeter off a normal one with ethanol wiped clean scissors or razor blade.

7. Incubate for 3 min (5 min for yeast cells), with gentle tapping at the incubation midpoint.

RNA extraction l TIMING: 1 h

8. Extract RNA using Option A for non-yeast nuclei or permeabilized cells or Option B for yeast.

A. RNA extraction from non-yeast nuclei or permeabilized cells

a. Add 500 μl Trizol LS and mix well by vortexing to stop the reaction.

b. Incubate the homogenized sample for 5 min at room temperature (25^0C) to allow the complete dissociation of nucleoprotein complexes and add 130 μl Chloroform.

c. Vortex sample vigorously for 15 s and incubate at room temperature for 2 to 3 min.

d. Centrifuge at 14,000 *g* for 5 min at 4°C, transfer the aqueous phase to a new tube, and add 1 μl GlycoBlue.

B. RNA extraction from yeast cells or nuclei

a. Pellet cells or nuclei after the run-on reaction at 400 g for 5 mins at 4 °C and quickly resuspend in 500 μl phenol.

b. Add an equal volume of AES buffer and incubate it at 65°C for 5 min with periodic vortexing. Let the mixture rest on ice for 5 min, and then add 200 μl of chloroform.

c. Vortex sample vigorously for 15 s and incubate at room temperature for 2 to 3 min.

d. Centrifuge at 14,000 *g* for 5 min at 4°C, transfer the aqueous phase to a new tube, and add 1 μl GlycoBlue and NaOAc to 200 mM final conc.

9. Add 2.5x volume of 100% room temperature ethanol & vortex for 10 s.
10. Incubate samples at room temperature for 10 min.
11. Centrifuge at 14,000 *g* for 20 min at 4°C. The RNA precipitate forms a gel-like pellet on the side and bottom of the tube.
12. Remove supernatant completely.
13. Add 750 μl of 75% ethanol.

    PAUSE POINT: The RNA pellet in 75% ethanol can be stored up to a week at -80°C.
14. Mix by vortexing and centrifuge at 14,000 *g* for 5 min at 4°C.
15. Remove all supernatant.
16. Air-dry the RNA pellet for 5-10 min.

    CRITICAL STEP: It is important not to let the RNA pellet dry completely as this will greatly decrease its solubility.
17. Re-dissolve the RNA pellet in 20 μl DEPC H_2O.

RNA fragmentation by base hydrolysis | Timing: 0.5 h

18. Heat denature the RNA at 65°C on a heat block for 40 s and then place the tubes on ice.
19. Add 5 μl of ice cold 1 N NaOH & incubate the mixture on ice for 10 min.
20. Neutralize with 25 μl of 1 M Tris-HCl pH 6.8.

21. Perform buffer exchange once by running the 50 μl base-hydrolyzed RNA sample through a P-30 column according to the manufacturer's instructions.
    a. Invert the column sharply several times to resuspend the settled gel and remove any bubbles.
    b. Snap off the tip and place column in a 2.0 ml microcentrifuge tube (included).
    c. Remove cap. Allow the excess packing buffer to drain by gravity to top of gel bed. (If column does not begin to flow, push cap back into column and remove).
    d. Discard the drained buffer, then place the column back into 2 ml tube.
    e. Centrifuge for 2 min at 1,000 x to remove the packing buffer.
    f. Discard the buffer.
    g. Place the column in a clean 1.5 ml microcentrifuge tube.
    h. Carefully apply the sample (20 - 100 μl) directly to the center of the column. Application of more or less than the recommended sample volume may decrease column performance.
    i. After loading sample, centrifuge the column for 4 min at 1,000 x g.
    j. Following centrifugation, the purified sample is now in Tris or SSC buffer. Molecules smaller than the column's exclusion limit (20 bp) will be retained.
22. Add 1μl RNase inhibitor to the elution (base-hydrolyzed nascent RNA).

First biotin RNA enrichment | Timing: 3 h

23. Prepare the buffers for bead binding and washing. The buffer recipe provided below would suffice for three bead binding and wash steps of ePRO-seq for 12 libraries.

Buffer	**NaCl (5 M)**	**NaOH (1 N)**	**DEPC H_2O**
NaOH Wash	50 μl (50 mM)	500 μl (0.1 N)	4.45 ml
NaCl Wash	200 μl (100 mM)	-	9.8 ml

Buffer	**Tris-Cl 7.4 (1 M)**	**NaCl (5 M)**	**Triton X-100 (10%)**	**DEPC H_2O**
High Salt Wash	1 ml (50mM)	8 ml (2M)	1 ml (0.5%)	10 mL
Binding Buffer	200μl (10mM)	1.2 ml (300mM)	200 μl (0.1%)	18.4 mL
Low Salt Wash	50μl (5mM)	-	100 μl (0.1%)	9.85 mL

24. Take 30 μl of Streptavidin coated magnetic beads per library.
25. Place the beads on the magnetic separator for 1 min and discard the supernatant.
26. Pre-wash the beads in 1 ml NaOH Wash buffer for 2 min.
27. Place on the magnetic separator for 2 min and remove supernatant.
28. Wash beads twice with 1 ml NaCl Wash buffer.
29. Resuspend the beads in 50 μl of the Binding Buffer per library.
30. Mix the RNA sample from step 22 with 50 μl of pre-washed Streptavidin beads.
31. Incubate at room temperature on a rotator set at 8 rpm for 20 min.

32. Place on magnet for 2 min and remove the liquid.
33. Resuspend the beads in 500 µl ice cold High Salt Wash.
34. Wash for 2 min using rotator.
35. Place on magnet for 2 min and remove the buffer.
36. Repeat High Salt Wash by following steps 33-35.
37. Wash twice with 500 µl Binding Buffer.
38. Wash once with 500 µl Low Salt Wash.
39. Resuspend the beads in 300 µl Trizol and vortex thoroughly.
40. Incubate for 3 min at room temperature.
41. Add 60 µl chloroform, vortex thoroughly for 20 s, and incubate for 3 min at room temperature
42. Centrifuge at 20,000 *g* for 5 min at 4°C.
43. Transfer the aqueous layer to a new tube.
44. Remove and discard the organic phase, leaving the beads and the unpipetted aqueous phase.
45. Extract RNA from the beads once more by repeating steps 38-41.
46. Collect the aqueous layer and combine with the sample from step 43.
47. Add 1 µl of GlycoBlue and 900 µl of 100% ethanol and vortex thoroughly.
48. Incubate samples at room temperature for 10 min.
49. Centrifuge at 20,000 *g* for 20 min at 4°C.
50. Discard the supernatant from the tube, leaving only the RNA pellet.
51. Wash the RNA pellet with 750 µl of 75% ethanol.

PAUSE POINT: The RNA pellet in 75% ethanol can be stored up to a week at -80°C.

52. Mix by vortexing and centrifuge at 20,000 *g* for 5 min at 4°C.

53. Discard the supernatant and remove all residual liquid.

54. Air-dry the RNA pellet for 5-10 min.

55. Dissolve the RNA pellet in 3.5 μl DEPC H_2O.

3' RNA adaptor ligation I Timing: 4.5 h

56. In ePRO-seq, the barcode for multiplexing is present in the 3' RNA adaptor PCRfreeRC3_BRn. Therefore, differently barcoded 3' RNA adaptors (PCRfreeRC3_BR1, PCRfreeRC3_BR2…) should be used for different libraries that will be pooled together for sequencing.

    The 5' adaptors are at 100 μM concentration in -80^{0}C (Lis Lab -80^{0}C I)

57. Add 0.5 μl of the appropriate 3' RNA adaptor into the 3.5 μl of RNA.

58. Mix gently, heat denature at 65°C for 20 s, then place on ice.

59. Make the RNA ligation mix tabulated below. When processing multiple samples, scale up accordingly.

Reagents	**Volume per reaction (μl)**	**Final concentration**
T4 RNA ligase buffer (10x)	1	1x
ATP (10 mM)	1	1 mM
50% PEG	2	10 %
RNase inhibitor	1	4 units per μl
T4 RNA ligase I	1	1 units per μl

60. Add 6 µl of the mix to the 4 µl of RNA (10 µl final).

    CRITICAL STEP: GlycoBlue may form a precipitate in the presence of high PEG concentration, but is not reported to affect the ligation efficiency.

61. Incubate at 20°C for 4 hr then place at 4°C until ready to proceed to the next step.

    PAUSE POINT: The ligation reaction can be left at 4°C overnight.

Second biotin RNA enrichment | Timing: 3 h

62. Bring the volume of the adaptor ligated RNA to 50 µl by adding 40 µl DEPC H_2O.

63. Perform a second biotin enrichment by repeating steps 24-32.

64. Resuspend the beads in 200 µl ice cold High Salt Wash and pool into two tubes.

    CRITICAL: Libraries with different in-line barcodes can be pooled at this step.

    Pooling into even number of tubes is recommended for easier handling and centrifuge balancing.

    I usually pool six libraries per tube and the reaction volumes hereon are optimized accordingly.

65. Wash the magnetic beads twice with High Salt Wash, twice with Binding Buffer, and once with Low Salt Wash following steps 34-38.

66. Trizol extract RNA following steps 39-54.

67. Re-dissolve the RNA pellet in 10 µl DEPC H_2O.

68. Heat denature briefly at 65°C for 20 s and place in ice.

Enzymatic modification of the RNA 5' ends l Timing: 3.5-4 h

69. Prepare 5' cap repair enzyme mix: depending on the availability of Tobacco acid pyrophosphatase (TAP) or RNA 5' Pyrophosphohydrolase (RppH). When using TAP, use the first table. When using RppH, prepare the enzyme mix in the second table. When processing multiple samples, scale up accordingly.

Reagents	**Volume per reaction (µl)**	**Final concentration**
DEPC H_2O	6	
TAP buffer (10x)	2	1x
RNase inhibitor	1	2 units per µl
TAP	1	0.5 units per µl

Reagents	**Volume per reaction (µl)**	**Final concentration**
DEPC H_2O	5	
ThermoPol Reaction buffer (10x)	2	1x
RNase inhibitor	1	2 units per µl
RppH	2	0.5 units per µl

70. Add 10 µl of the appropriate enzyme mix to the 10 µl of RNA from step 68.

71. Incubate at 37°C for 1 hr.

Hydroxyl repair | TIMING: 2 h

72. Prepare PNK mix as tabulated below. When processing multiple samples, scale up accordingly.

Reagents	**Volume per reaction (μl)**	**Final concentration**
DEPC H_2O	55	
PNK buffer (10x)	10	1x
ATP (10 mM)	10	1 mM
RNase inhibitor	2.5	1 units per μl
Polynucleotide kinase (PNK)	2.5	0.25 units per μl

73. Add 80 μl of the mix to the 20 μl of RNA from step 71.

74. Incubate at 37°C for 1 hr.

75. Add 300 μl of Trizol and vortex for 5 s.

76. Add 60 μl chloroform, vortex for 15 s, and incubate for 2 min at room temperature.

77. Centrifuge at 20,000 *g* for 5 min at 4°C.

78. Transfer aqueous layer to a clean microfuge tube.

79. Add 1 μl GlycoBlue and 2.5x volume of 100% ethanol.

80. Pellet the RNA by centrifuging at 20,000 *g* for 20 min at 4°C.

81. Discard the supernatant from the tube, leaving only the RNA pellet.

82. Wash the RNA pellet in 75% ethanol.

    PAUSE POINT: The RNA pellet in 75% ethanol can be stored up to a week at -80°C.

83. Mix by vortexing and centrifuge at 20,000 *g* for 5 min at 4°C.

84. Discard the supernatant and remove all residual liquid.

85. Air-dry the RNA pellet for 5-10 min.

86. Dissolve the RNA pellet in 6 μl DEPC H_2O.

5' RNA adaptor ligation I Timing: 4.5 h

87. ePRO-seq uses PCRfreeRC5 for single end sequencing and 5Adapt_Paired_ePRO for paired end sequencing. The 5' adaptors are at 100 μM concentration in -80^{0}C (Lis Lab -80^{0}C I)

88. Add 2 μl of the 100 μM 5' RNA adaptor.

89. Mix gently, heat denature at 65°C for 20 s, then place on ice.

90. Make the RNA ligation mix as described below.

Reagents	**Volume per pool (μl)**	**Final concentration**
T4 RNA ligase buffer (10x)	2	1x
ATP (10 mM)	2	1 mM
50% PEG	4	10 %
RNase inhibitor	1	2 units per μl
T4 RNA ligase I	3	1.5 units per μl

91. Add 12 μl of the RNA ligation mix to the 8 μl of RNA (20 μl final).

CRITICAL STEP: GlycoBlue may form precipitate in the presence of high PEG concentration, but is not reported to affect the ligation.

92. Incubate at 20°C for 4 hr then place at 4°C until ready to proceed to the next step.

PAUSE POINT: The ligation reaction can be left at 4°C overnight.

Third biotin RNA enrichment | Timing: 3 hr

93. Bring the volume of the adaptor ligated RNA to 50 µl with 30 µl DEPC H_2O.

94. Take 90 µl of Streptavidin coated magnetic beads per pool.

95. Perform third biotin enrichment of RNA by repeating steps 25-54.

96. Dissolve the RNA pellet in 10 µl DEPC H_2O.

Reverse transcription | Timing: 2 h

97. Make reverse transcription (RT) primer mix as shown below.

Reagents	**Amount per pool (µl)**	**Final concentration**
RP1 reverse transcription primer (100 µM)	2	8 uM
dNTP mix (12.5 mM)	2	1 mM

98. Add 4 µl of the RT primer mix to the 10 µl of re-dissolved RNA.

99. Heat to 65°C for 5 min, chill on ice for 2 min, and briefly spin at 1,000g for 30 s.

100. Prepare the RT reaction mix as tabulated below.

When processing multiple samples, scale up accordingly.

Reagents	**Volume per pool (µl)**	**Final concentration**
First strand buffer (5x)	5	1x
DTT (0.1 M)	1.5	6 mM
RNase inhibitor	1.5	1 units per µl
Superscript III	3	24 U/µl

101. Add 11 µl of the RT reaction mix to the 14 ul of RNA-primer mix and incubate for 5 min at RT.

102. Transfer the reaction to 0.5 ml tubes.

103. Reverse transcribe the RNA in PCR block for 15 min at 45°C, 40 min at 50°C, 10 min at 55°C, and 15 min at 70°C.

PAUSE POINT: The reverse transcribed cDNA can be stored for a month at -20°C.

RNA degradation and cDNA purification l Timing: 2 h

104. Degradation of RNA and purification of cDNA is critical in ePRO-seq because the cDNA is directly used for sequencing without further manipulation. Prepare RNase reaction mix as tabulated below.

Reagents	**Volume per pool (µl)**	**Final concentration**
RNase H reaction buffer (10x)	3.5	1x
RNase cocktail (0.5 U/µl of RNAse A, 20 U/µl of RNase H)	3	0.04 unit/µl of RNase A & 1.7 U/µl of RNase T1
RNase H (5 U/µl)	3.5	0.5 unit per µl

105. Add 10 ul of the RNase reaction mix to 25 ul of RT reaction.

106. Incubate the reaction for 20 min at 37^0C for RNA degradation.

107. Incubate another 20 min at 65^0C for heat inactivation of the enzymes.

108. Once the RNA digestion is complete, purify the cDNA by phenol chloroform extraction as below.

a. Add 215 µl DEPC H_2O to 35 µl of RNase reaction to bring the final volume to 250 µl.
b. Add an equal volume of buffered phenol:chloroform, and vortex thoroughly.
c. Centrifuge at 20,000 *g* for 5 min at 4°C.
d. Collect the aqueous layer in a clean tube.
e. Follow steps 79-85.

109. Re-dissolve the pellet in appropriate volume.

For example, when making 12 libraries, pooling 6 libraries each in two tubes, resuspend cDNA in each tube in 20 µl DEPC H_2O. Of the 40 µl after combining the cDNA from two tubes, keep 20 µl as backup and send 20 µl to sequencing facility.

CRITICAL: The 20 µl backup cDNA can be PCR amplified using RP1 and TestAmp3 primers if the amount of ePRO-seq library is not sufficient for sequencing.

110. Quantification of the library is performed by the facility using digital PCR. Request the sequencing facility to use TRU-seq small RNA primers for digital PCR.

Test PCR amplification | Timing: 2 h

Eliminated in ePRO-seq.

However, it can be done using RP1 and TestAmp3 primers. Follow PRO-seq protocol for steps.

Gel analysis of test PCR | Timing: 2 h

Eliminated in ePRO-seq.

However, if Test PCR amplification is performed, gel analysis of test PCR can be performed following PRO-seq protocol.

Full-scale PCR amplification | Timing: 2.5 h

Eliminated in ePRO-seq.

However, if the library concentration measured using digital PCR is insufficient for sequencing, libraries can be amplified usig RP1 and TestAmp3 primers.

Library size selection by PAGE | Timing: 5 h to 1 day

Eliminated in ePRO-seq.

If full-scale amplification is performed, libraries should be size-selected following steps in PRO-seq protocol.

High-throughput sequencing | Timing: 24 h

111. Sequencing of ePRO-seq requires slight modification. Since the library is single-stranded DNA, quantification by intercalating dye-based methods such as Qubit may not be accurate. We recommend digital PCR for quantification of cluster-generating DNA molecules in the library. Many sequencing facilities provide this service with a small additional cost (~$20). Once the library concentration is quantified, we recommend loading twice the concentration of normal double-stranded library. For example, the concentration of PRO-seq dsDNA library loaded into Illumina NextSeq500 is 2 nM and we recommend loading 4 nM of ePRO-seq library.

Data analysis | TIMING: variable

In both ePRO-seq & PRO-seq, the modified RNA adaptors were designed to enable sequencing of the reverse complement of nascent RNAs. Therefore, the 3' end of the reverse complement of the sequencing reads reflects the RNA polymerase active site position. Below, we outline the three major stages of a simple processing pipeline.

112. Pre-process the raw sequence data. Filter out low quality reads. Separate the barcoded reads based on the first six bp of the sequences. Once individual libraries are separated, trim the first seven bps of the reads (six bp barcode + 1 one G that was a part of 3' adaptor) and clip

the potential adaptor sequences (ATCTCGTATGCCGTCTTCTGCTTG for single-end sequencing) from the sequence reads. Tools such as 'fastx_trimmer' and 'fastx_clipper' are publicly available for this purpose. Depending on the quality of the library, sequences containing only the adaptor sequences (adaptor dimers) may take up to 5% of total reads. Once the sequences are cleaned up, make reverse complements of sequences and the sequences are ready for mapping.

113. Map or align the sequence reads to the genomic sequence. Since most of the nascent RNA reads are captured before RNA processing and splicing, they do not contain large gaps in alignment. Therefore, many alignment programs based on the Burrows-Wheeler transformation algorithm such as 'bwa' or 'bowtie' work well. Reads with multiple alignments are usually discarded, unless they are used for studying specific target regions that are repeated more than once. On average, about 55-70% of the raw sequence reads are aligned uniquely to the genome. The alignment results are commonly stored in 'sam' or 'bam' formats.

114. Generate the coverage of the aligned sequence reads. A common way to do this using publicly available tools is as follows: first, sort the 'bam' file using 'samtools sort'; then process the sorted 'bam' file using 'bedtools genomecov' with '-ibam' (use bam file input), '-strand' (strand specific coverage), and '-5' (5' position coverage') options. For the PRO-seq

data, swap the plus and minus strand data for the correct orientation. These data can be visualized in genome browsers (Figure A.3), and used in further downstream analyses.

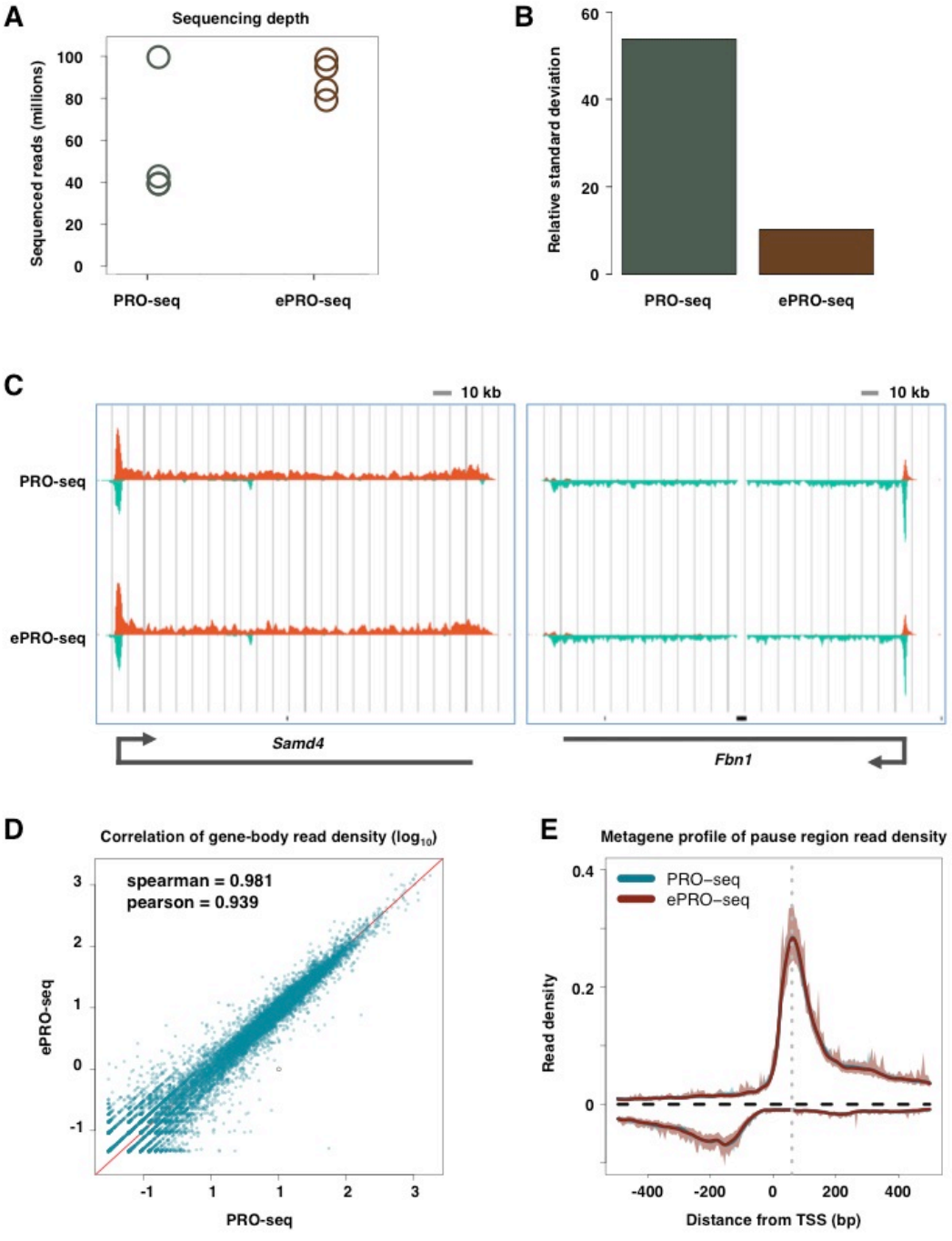
Anticipated results

The final product of the ePRO-seq is genome-wide map of RNA polymerase active sites. In general, ePRO-seq and PRO-seq profile of a gene exhibits several features such as: a) higher read density at the promoter-proximal pause site compared to the gene body representing paused Pol IIs, b) uniform read density across exons and introns, c) reads beyond the polyadenylation site, and d) divergent PRO-seq reads at the promoter of genes in mammals indicating divergent transcription. The enhancer regions, which are present in intergenic and intragenic regions, are also characterized by divergent PRO-seq reads. Spiky PRO-seq read coverage along the gene body indicates lower complexity of the library, which may arise from use of fewer nuclei or permeabilized.

ePRO-seq enables pooling of differentially barcoded libraries during or after the second biotin RNA enrichment. This ability to pool libraries earlier during the library preparation simplifies handling and allows for multiplexing. In addition to these advantages, we noticed that the pooled libraries have lower variance in sequencing depth compared to PRO-seq libraries that are pooled just prior to sequencing (Figure B.2 A). The relative standard deviation of

sequencing depth of pooled libraries is lower in ePRO-seq compared to PRO-seq (Figure B.2B). However, the library quality and the transcription profile of genes are very similar between the two assays (Figure B.2C), suggesting the ePRO-seq library quality is just as good as PRO-seq. The similarity in library quality between ePRO-seq and PRO-seq is also evident at genome-wide level as shown be high correlation of gene-body reads (Figure B.2D) and promoter-proximal paused peaks (Figure B.2E). Overall, these preliminary analyses indicate that library generated using ePRO-seq is just as good as PRO-seq library, despite the shorter and easier library preparation procedures. The elimination of PCR- amplification of library and size-selection should remove the biases associated with them, at least in theory. Nevertheless, we did not observe change in library quality between ePRO-seq and PRO-seq, at least at the global level. This lack of difference could be explained either by the lack of such PCR and size-selection biases in PRO-seq or perhaps is more pronounced in high-depth libraries.

Figure B.2. Comparison of ePRO-seq with PRO-seq
(A) Sequencing depth variation among pooled libraries in PRO-seq (individually prepared and pooled prior to submission for sequencing in Illumina NextSeq500 lane) and ePRO-seq (pooled during the second biotin RNA enrichment step on the 2nd day of library preparation and sequenced in Illumina NextSeq500 lane).
(B) Relative standard deviation of sequencing depth of pooled libraries in PRO-seq and ePRO-seq.
(C) Screenshot of two representative genes (*Samd4* is in sense strand and *Fbn1* is in anti-sense strand) showing PRO-seq and ePRO-seq profiles.
(D) Correlation of gene body reads between PRO-seq and ePRO-seq.
(E) Metagene profiles of PRO-seq and ePRO-seq reads 500 bp around the transcription start sites (TSS).


A
Sequencing depth
Sequenced reads (millions)
100
80
60
40
20
0
PRO-seq
ePRO-seq
B
Relative standard deviation
60
40
20
0
PRO-seq
ePRO-seq
C
10 kb
10 kb
PRO-seq
ePRO-seq
Samd4
Fbn1
D
Correlation of gene-body read density (log10)
spearman = 0.981
pearson = 0.939
ePRO-seq
PRO-seq
3
2
1
0
-1
-1
1
1
2
3
E
Metagene profile of pause region read density
PRO−seq
ePRO−seq
Read density
0.4
0.2
0
-400
-200
0
200
400
Distance from TSS (bp)

Timing

Cell culture: 24 h

Sample preparation (nuclei isolation/cell permeabilization): 1 h

Nuclear Run-On: 1.5 h

RNA extraction: 1 h

RNA fragmentation by base hydrolysis: 0.5 h

Biotin RNA enrichment: 3 h

3’ RNA adaptor ligation: 0.5 h hands-on and 4 h incubation; can be stored overnight (suggested end of the first day)

Second biotin RNA enrichment: 3 h

Enzymatic modification of the RNA 5′ ends: 3.5–4 h

5’ RNA adaptor ligation: 0.5 h hands-on and 4 h incubation; can be stored overnight (suggested end of the second day)

Third biotin RNA enrichment: 3 h

Reverse transcription: 2 h

RNA degradation and cDNA purification: 2 h

High-throughput sequencing, 24 h

Data analysis: variable

REFERENCES

Aird, D., Ross, M.G., Chen, W.-S., Danielsson, M., Fennell, T., Russ, C., Jaffe, D.B., Nusbaum, C., and Gnirke, A. (2011). Analyzing and minimizing PCR amplification bias in Illumina sequencing libraries. Genome Biol. *12*, R18.

Collart, M.A., and Oliviero, S. (2001). Preparation of Yeast RNA (Hoboken, NJ, USA: John Wiley & Sons, Inc.).

Core, L.J., Martins, A.L., Danko, C.G., Waters, C.T., Siepel, A., and Lis, J.T. (2014). Analysis of nascent RNA identifies a unified architecture of initiation regions at mammalian promoters and enhancers. Nat Genet *46*, 1311–1320.

García-Martínez, J., Aranda, A., and Pérez-Ortín, J.E. (2004). Genomic run-on evaluates transcription rates for all yeast genes and identifies gene regulatory mechanisms. Molecular Cell *15*, 303–313.

Job, D., Durand, R., Job, C., and Teissere, M. (1984). Complex RNA chain elongation kinetics by wheat germ RNA polymerase II. Nucleic Acids Research *12*, 3303–3319.

Jones, M.B., Highlander, S.K., Anderson, E.L., Li, W., Dayrit, M., Klitgord, N., Fabani, M.M., Seguritan, V., Green, J., Pride, D.T., et al. (2015). Library preparation methodology can influence genomic and functional predictions in human microbiome research. Proceedings of the National Academy of Sciences *112*, 14024–14029.

Kwak, H., Fuda, N.J., Core, L.J., and Lis, J.T. (2013). Precise maps of RNA polymerase reveal how promoters direct initiation and pausing. Science *339*, 950–953.

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol.

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with

Burrows-Wheeler transform. Bioinformatics *25*, 1754–1760.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078–2079.

Mahat, D.B., Salamanca, H.H., Duarte, F.M., Danko, C.G., and Lis, J.T. (2016a). Mammalian Heat Shock Response and Mechanisms Underlying Its Genome-wide Transcriptional Regulation. Molecular Cell *62*, 63–78.

Mahat, D.B., Kwak, H., Booth, G.T., Jonkers, I.H., Danko, C.G., Patel, R.K., Waters, C.T., Munson, K., Core, L.J., and Lis, J.T. (2016b). Base-pair-resolution genome-wide mapping of active RNA polymerases using precision nuclear run-on (PRO-seq). Nat Protoc *11*, 1455–1476.

Poptsova, M.S., Il'icheva, I.A., Nechipurenko, D.Y., Panchenko, L.A., Khodikov, M.V., Oparina, N.Y., Polozov, R.V., Nechipurenko, Y.D., and Grokhovsky, S.L. (2014). Non-random DNA fragmentation in next-generation sequencing. Sci Rep *4*, 4532.

Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics *26*, 841–842.

APPENDIX C[f]

CBP FACILITATES TRANSCRIPTION ELONGATION BY ENABLING RNA POLYMERASE TO TRAVERSE +1 NUCLEOSOME BARRIER

Summary

Transcription activation involves RNA polymerase II (Pol II) recruitment and release from the promoter into productive elongation, but how specific chromatin regulators control these steps is not fully understood. Here we identify a novel activity of the co-regulator and histone acetyltransferase p300/CBP in controlling promoter-proximal paused Pol II. Inhibition of CBP acetyl-transferase activity rapidly reduces Pol II recruitment to the highly paused promoters. CBP inhibition impedes transcription through the +1 nucleosome causing Pol II to accumulate at this position on highly or moderately paused promoters. Overall, we uncover two key roles of CBP in directly stimulating Pol II recruitment and enabling Pol II to traverse the first nucleosome.

[f] Adapted from Boija, A.*, Mahat, D. B.*, Zare, A., Holmqvist, P.H., Philip, P., Meyers, D.J., Cole, P.A., Lis, J.T., Stenberg, P., & Mannervik, M. (2016). CBP regulates promoter-proximal RNA polymerase II. *In review*. *denotes equal contribution. Contributions are shown in each figure.

Introduction

Regulation of transcription occurs at two major steps in metazoans (reviewed by Core and Lis, 2008). The first step is the recruitment of RNA Polymerase II (Pol II) to the promoter, which is orchestrated by sequence-specific transcription factors that enable formation of a pre-initiation complex (PIC), consisting of Pol II and general transcription factors (GTFs). A successful formation of the PIC leads to rapid initiation and transcription of 20 to 60 nucleotides (Core et al., 2012; Venters and Pugh, 2016). This brings Pol II to the second step of regulation where it pauses on many genes, an event mediated by the action of NELF and DSIF. The release of paused Pol II into productive elongation requires recruitment of P-TEFb kinase, which phosphorylates these pausing factors and the Pol II C-terminal domain (CTD) (reviewed in Zhou et al., 2012). Although pausing temporarily restrains Pol II from entering elongation, a majority of paused genes are highly expressed (reviewed in Adelman and Lis, 2012). In most instances, pausing may therefore be a mechanism to maintain or tune expression rather than to function as an on-off switch.

Our understanding of how specific chromatin regulators are recruited to promoters and how they regulate the two major steps in transcription is incomplete. CREB-binding protein (CBP) and its paralog p300 are widely used transcriptional co-regulators with over 400 interaction partners (reviewed in Bedford et al., 2010). p300/CBP has a histone acetyltransferase (HAT) activity

and is known to acetylate lysine 18 and 27 of histone H3 and lysine 8 of histone H4, establishing an active chromatin state (Feller et al., 2015; Jin et al., 2011; Tie et al., 2009). However, the mechanisms by which p300/CBP regulates transcription are not fully understood.

Herein, we analyzed the functions of the sole p300/CBP homolog (*nejire*) at gene promoters in *Drosophila* S2 cells. Run-on (PRO-seq) experiments following CBP inhibition demonstrated that CBP functions to position Pol II at the pause site, and facilitates efficient release of paused Pol II into productive elongation by overcoming the transcriptional barrier caused by the +1 nucleosome. CBP is also important for efficient Pol II recruitment to promoters. We suggest that CBP activity is rate-limiting for Pol II release from weakly-paused genes, and rate-limiting for recruitment to the highly-paused promoters.

Materials and methods

Drug treatment and nuclei isolation of S2 cells

Nuclei were isolated as described previously (Core et al., 2013). In brief, 10 ml of 2 x 10^6 cells/ml of S2 cells were treated with CBP inhibitor (C646) or control drug (C37) in DMSO for 10 min followed by a PBS wash. Cells were resuspended in buffer L (10 mM Tris-HCl pH 7.5, 300mM sucrose, 10mM NaCl, 3 mM $CaCl_2$, 2 mM $MgCl_2$, 0.1% Triton X, 0.5 mM DTT, protease inhibitors cocktail (Roche), 4 u/ml RNase inhibitor (SUPERaseIN, Ambion) and

immediately dounced 25 strokes with a tight fitting pestle. Lysed cells were mixed with an equal amount of buffer B (10 mM Tris-HCl pH 8, 2M sucrose, 10mM $NaCl_2$, 2 mM $MgCl_2$, 0.5 mM DTT, protease inhibitors cocktail (Roche), 4 u/ml RNase inhibitor (SUPERaseIN, Ambion) and loaded onto a buffer B sucrose pillow. The sample was spun at 23000g on a SW-41 rotor, the supernatant was removed, and the nuclei were washed once in storage buffer (50 mM Tris-HCl pH 8, 25% glycerol, 5mM $MgAc_2$, 0.1mM EDTA, 5mM DTT) and recovered by centrifuging at 1000g for 5 minutes. Isolated nuclei were resuspended in storage buffer and kept at -80^{0}C. Nuclei from two biological replicates were prepared for each condition.

Nuclear run-on and PRO-seq library preparation

Nuclear run-on was carried out as described previously with some modifications (Kwak et al., 2013). Briefly, 10 million nuclei in 100 µl of storage buffer were mixed with 100 µl of 2x nuclear run-on buffer (10 mM Tris-HCl pH 8.0, 5 mM $MgCl_2$, 1 mM DTT, 300 mM KCl, 1% Sarkosyl, 50 uM biotin-11-A/C/G/UTP, 0.2 units/µl RNase inhibitor) and incubated at 30^{0}C for three minutes. RNAs were isolated using TRIZOL LS and base-hydrolyzed with 200 nM final concentration of NaOH, generating an average size of RNA between 100-150 nucleotides. Nascent RNAs from run-on reactions, characterized by the addition of biotinylated nucleotide, were isolated with magnetic beads coated with streptavidin. After the ligation of 3' adapter and a second biotin

streptavidin affinity purification, the mRNA cap was removed and the 5' adapter was ligated. Then a third biotin streptavidin affinity purification was carried out and cDNA was generated by reverse transcription. The generated cDNA was amplified with 9 cycles of PCR using Illumina TruSeq small-RNA adaptors for sequencing.

Normalization

To normalize the datasets from the two treatments we could not apply standard normalization techniques as they assume that only a minority of the data are changed between treatments (Landfors et al., 2011). Since CBP occupancy is higher than the genomic mean at virtually all expressed promoters (data not shown), we expect it to affect expression of a large number of genes. We therefore normalized the PRO-seq datasets using the TSS proximal signal (TSS +100 bp) of the 5% (n=270) expressed genes with the least CBP signal. At these genes, we expect the smallest changes after CBP inhibition. After this normalization, we quantified the change in PRO-seq density upon CBP inhibition at all expressed genes.

Differential expression analyses of pause region and gene body

Significant change in PRO-seq density in the pause region and the body of gene was calculated using DESeq2 (Love et al., 2014) by analyzing the two biological replicates separately. PRO-seq count reads were calculated

from the 100 bp pause region or gene body (500 bp downstream of TSS to 100 bp upstream of poly A site). Adjusted p-value cutoff of 0.001 and fold change of at least 1.25 was used in DESeq2. For all other analyzes of the PRO-seq data the reads from the two biological replicates were merged.

Results

CBP is required for transcription of thousands of genes

To investigate CBP's functions in transcription, we used a previously described inhibitor C646 that selectively inhibits the catalytic activity of CBP (Bowers et al., 2010). As a control, we used C37, a compound very similar in structure to C646 but that shows no effect on CBP HAT activity (Bowers et al., 2010). C646 and related inhibitors have previously been shown to affect histone acetylation in *Drosophila* and mammalian cells within minutes (Crump et al., 2011; Dancy et al., 2012). To assess the immediate transcriptional response to CBP inhibition genome-wide, we performed precision run-on sequencing (PRO-seq) on S2 cells treated with C37 (control) or C646 for 10 min (Figure C.1A). This genomic assay maps transcriptionally engaged Pol II with single nucleotide resolution (Kwak et al., 2013). The biological replicates of PRO-seq showed strong correlation at both pause region and gene body (Figure C.1B). After normalization, we quantified the change in PRO-seq density upon CBP inhibition at all expressed genes. We found 3790 genes with significant ($p < 0.001$) down-regulation but only 42 genes with significant

upregulation in gene-body PRO-seq density in C646-treated cells (Figure C.1C). However, all 42 up-regulated genes are false-positives – because the genes are very short (Figure C.1D), the PRO-seq density increase in the promoter-proximal region was attributed to the gene-body. We therefore conclude that the only direct effect of CBP inhibition is transcription down-regulation, and a large fraction of expressed genes in S2 cells require CBP for their transcription.

CBP inhibition reduces Pol II occupancy at highly-paused promoters

Global change in gene-body PRO-seq density prompted us to look at change in Pol II occupancy at promoters genome-wide. We measured PRO-seq density in the promoter-proximal region (TSS to 300 bp downstream of TSS) of all genes that showed significant down-regulation in gene-body (n=3790). This analysis identified two different responses to CBP inhibition - an increase in promoter-proximal Pol II at 2529 genes and a decrease in Pol II at 1243 genes (Figure C.2A). Interestingly, genes with increased promoter-proximal Pol II upon CBP inhibition have lower Pol II occupancy (weakly paused) in control cells than genes with decreased promoter-proximal Pol II (highly paused) (compare Figures C.2B with C.2C). This result indicates that genes with weakly paused promoters accumulate Pol II in the promoter-proximal region after CBP inhibition indicating that CBP has a function in

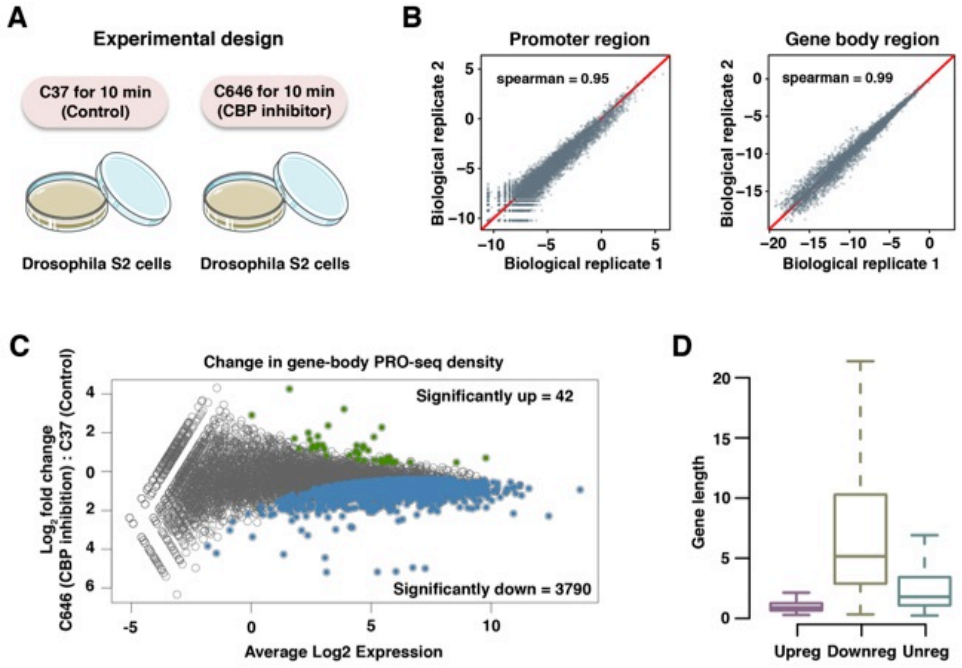


Figure C.1. CBP inhibition affects global transcription
(A) Schematics of PRO-seq experiment with CBP-inhibitor (C646) and control drug (C37).
(B) Correlation between biological replicates of PRO-seq reads in pause region (+1-100 bp) (left panel) and gene body region (500 bp from TSS to 100 bp upstream of poly A site) (right panel) in C646-treated S2 cells.
(C) Difference in gene-body precision run-on sequencing (PRO-seq) reads between 10 min CBP-inhibitor (C646) and control drug (C37)-treated S2 cells after normalization. Significantly up-regulated genes are indicated in green and significantly down-regulated genes are indicated in blue.
(D) Box plot of gene length for up-regulated with high confidence (Upreg), down-regulated with high confidence (Downreg), and genes without a significant change (Unreg) in gene body PRO-seq reads after CBP inhibition.

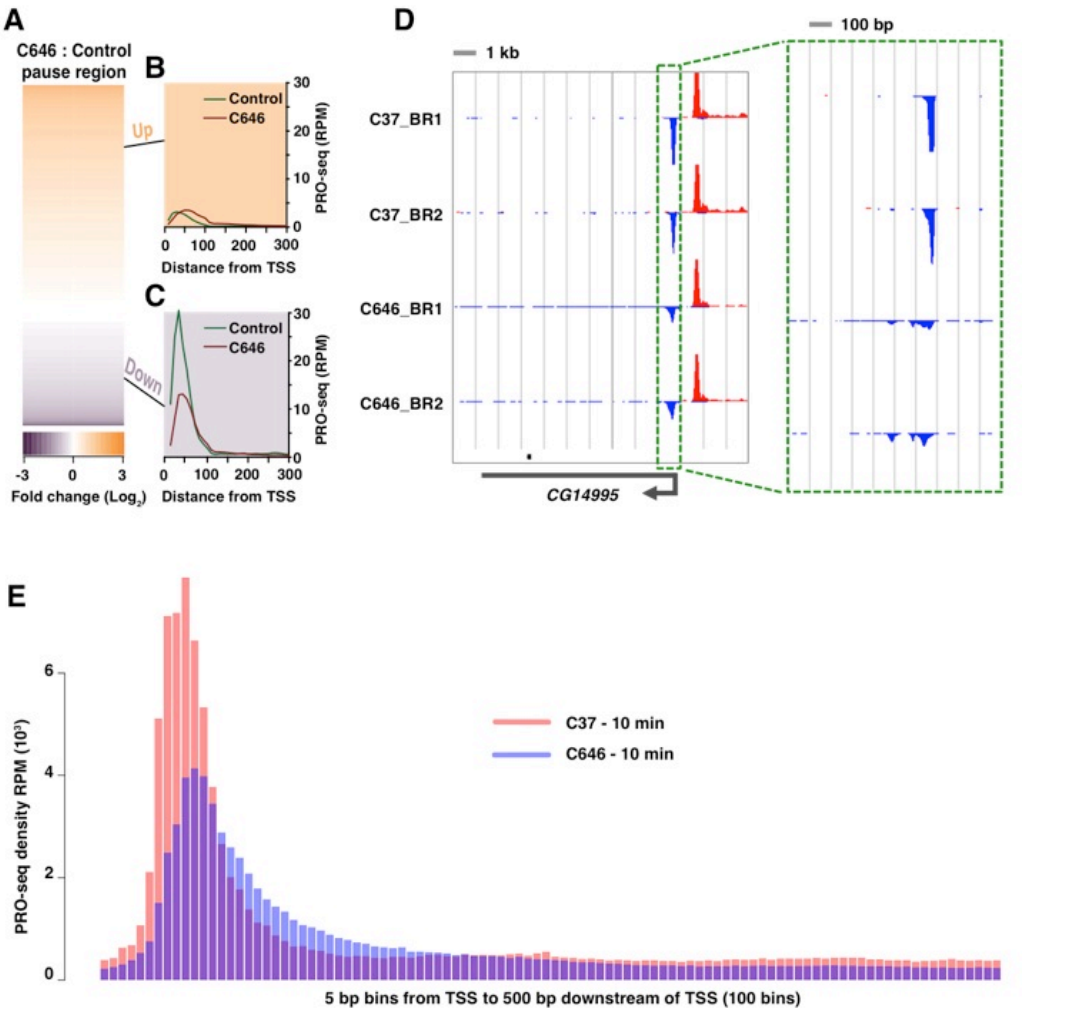


Figure C.2. CBP regulates Pol II positioning at promoter-proximal region
(A) Heatmap of PRO-seq fold change (FC) after 10 min CBP inhibition in the promoter-proximal region (TSS-300 bp) of genes (n=3790) with reduced gene body PRO-seq reads after 10 min of CBP inhibition (C646). 2529 genes had more PRO-seq reads in this region (shown in orange) and 1243 genes had less PRO-seq reads (shown in purple).
(B) PRO-seq metagene profiles for genes with decreased gene-body transcription after CBP inhibition but with increased (orange) promoter-proximal PRO-seq reads.
(C) PRO-seq metagene profiles for genes with decreased gene-body transcription after CBP inhibition but with decreased (purple) promoter-proximal PRO-seq reads.
(D) Screenshot of a gene showing PRO-seq density across the gene body in control drug and CBP-inhibitor treated cells. The inset elaborates the effect of CBP-inhibition in dribbling of promoter-proximal paused Pol II.
(E) Metagene profiles of PRO-seq reads in all expressed genes from TSS to 500 bp downstream in control or CBP-inhibited S2 cells for 10 min.

releasing Pol II into productive elongation. At highly paused genes, Pol II levels are decreased both in the promoter-proximal region and in the gene body, indicating that CBP has a crucial role in recruiting Pol II to these promoters.

CBP positions paused Pol II at the promoter-proximal site

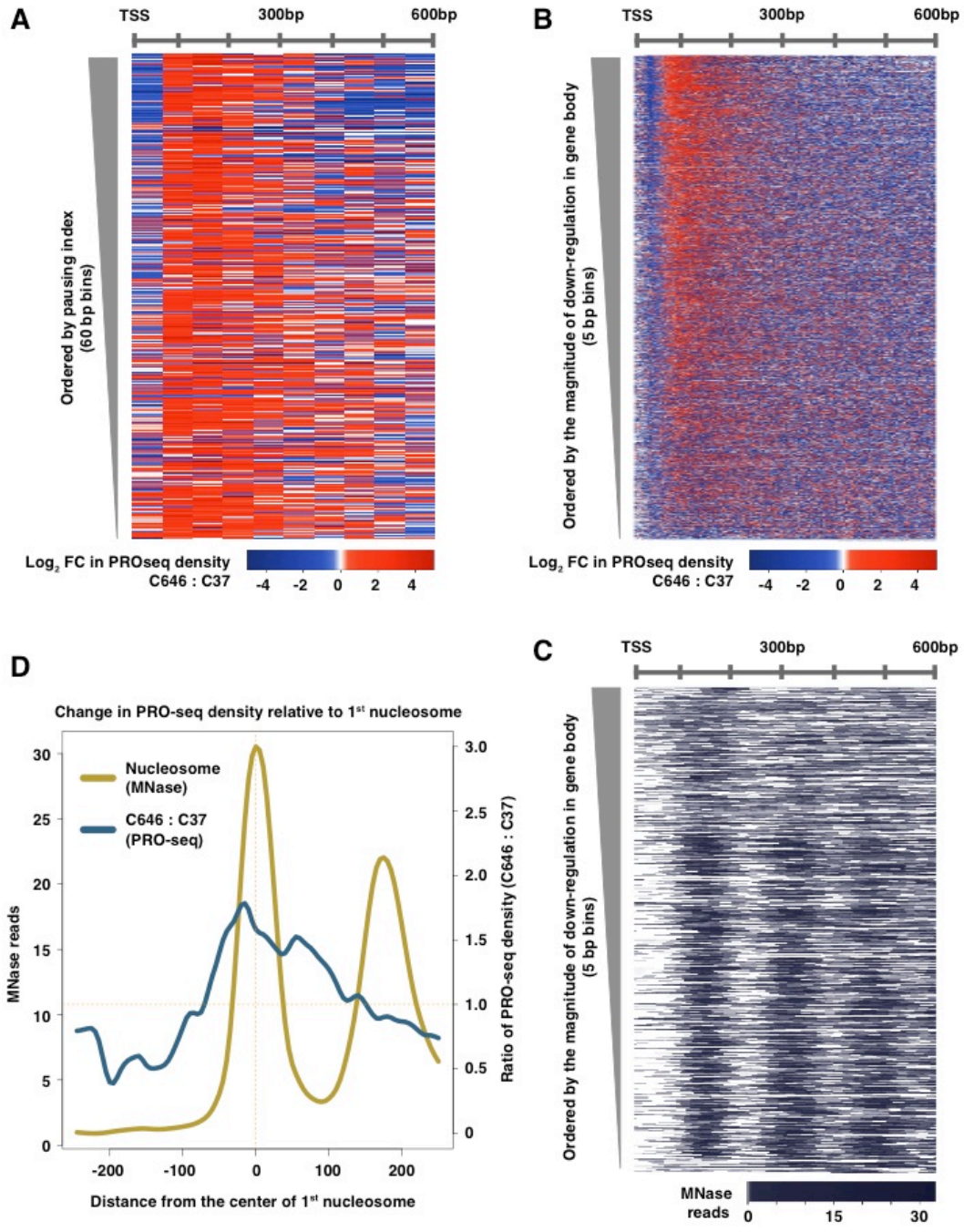
CBP inhibition not only affects the amount of Pol II at the promoter-proximal pause region but also alters the positioning of paused Pol II. We noted a shift in the position of transcriptionally engaged Pol II at both weakly-paused and highly-paused promoters upon CBP inhibition (see Figures C.2B and C.2C). Upon closer inspection at paused sites, we observed an accumulation of Pol II downstream of the canonical pause site in CBP-inhibited condition (Figure C.2D). A metagene profile of PRO-seq reads in the promoter-proximal region showed reduced level of Pol II at the canonical pause site, but Pol II "dribbles" downstream of the pause site after CBP inhibition (Figure C.2E). This observation indicates that CBP is important for proper positioning of Pol II at the canonical promote-proximal paused site.

CBP facilitates Pol II elongation by overcoming the +1 nucleosome barrier

To examine the scope of CBP in genome-wide positioning of paused Pol II, we measured the change in PRO-seq reads in 60 bp windows from the

TSS to 600 bp downstream in all expressed genes and ordered them by pausing index (Figure C.3A). The highly paused genes show a decrease in PRO-seq signal at the canonical pause site (0-60bp downstream of the TSS), whereas the less paused genes do not change as much. However, essentially all genes exhibit Pol II dribbling into the +60-180 bp region from TSS. Since this effect is observed after just 10 min of inhibitor treatment, it likely is a direct effect of CBP inhibition. The +60-180 bp region downstream of TSS where dribbled Pol II accumulate coincide with the position of first nucleosome. In order to examine the role of nucleosome in Pol II dribbling upon CBP inhibition, we plotted nucleosome density map from a previously published MNase-seq data (Gilchrist et al., 2010). The resolution of PRO-seq density change after CBP inhibition was increased to 5 bp bins, and the genes were ordered by the level of gene-body down-regulation. Interestingly, the increased PRO-seq density upon CBP inhibition correlates well with the position of the +1 nucleosome (Figures C.3B and C.3C), indicating that Pol II dribbles from the pause site but has difficulty overcoming the transcriptional barrier caused by the first nucleosome when CBP is inhibited. We then plotted the PRO-seq ratio in CBP inhibited versus control cells as a function of distance from the center of first nucleosome in control cells (Figure C.3D). This shows that the PRO-seq density in CBP inhibited cells is lower than in control cells in the pause region (around 100-250 bp upstream of the +1 nucleosome), but then increases until it reaches a maximum just upstream of the dyad axis of the +1

Figure C.3. **Promoter-proximal Pol II dribbles to more downstream positions after CBP inhibition but is retarded at the +1 nucleosome**
(A) Heatmap of log_2 fold change in PRO-seq reads in 60bp bins from the TSS to +600bp sorted on pausing index. The genes are ordered by their pausing index from top to bottom in decreasing order.
(B) Heatmap of log_2 fold change in PRO-seq reads in 5bp bins from the TSS to +600bp sorted on level of down-regulation in gene-body after CBP inhibition.
(C) Heatmap of MNase-seq data (Gilchrist et al., 2010) in 5bp bins from the TSS to +600bp sorted on level of down-regulation in gene-body PRO-seq desity.
(D) The ratio of PRO-seq change in CBP inhibited versus control cells and average profile of nucleosome position in control cells plotted as distance from the dyad axis of the +1 nucleosome.


A
TSS
300bp
600bp
Ordered by pausing index
(60 bp bins)
Log2 FC in PROseq density
C646 : C37
-4 -2 0 2 4
B
TSS
300bp
600bp
Ordered by the magnitude of down-regulation in gene body
(5 bp bins)
Log2 FC in PROseq density
C646 : C37
-4 -2 0 2 4
C
TSS
300bp
600bp
Ordered by the magnitude of down-regulation in gene body
(5 bp bins)
MNase
reads
0 15 30
D
Change in PRO-seq density relative to 1st nucleosome
Nucleosome
(MNase)
C646 : C37
(PRO-seq)
MNase reads
Ratio of PRO-seq density (C646 : C37)
0 5 10 15 20 25 30
0 0.5 1.0 1.5 2.0 2.5 3.0
-200 -100 0 100 200
Distance from the center of 1st nucleosome

nucleosome, from where it decreases again. These results further indicate that the Pol II is released from the canonical pause site, but cannot efficiently traverse the +1 nucleosome in absence of CBP.

CBP-mediated nucleosome modification facilitates Pol II elongation

Since CBP is a HAT that targets H4K8, H3K18 and H3K27 (Feller et al., 2015; Jin et al., 2011; Tie et al., 2009), we investigated if the inhibition of CBP affects the modifying marks and occupancy of +1 nucleosome. The level of H3K27ac in the promoter-proximal region, normalized by total H3 occupancy, was decreased upon CBP inhibition at both highly-paused and other promoters (Figure C.4A). In the first minute of treatment, the change in H3K27ac mark is very subtle if any, but a substantial reduction in H3K27ac is observed after 10 min of CBP inhibition. However, total H3 ChIP showed an opposite trend at all tested promoters - the overall nucleosome occupancy increased upon CBP drug treatment (Figure C.4B), suggesting a compaction of the chromatin. These results show that the chromatin compaction in the promoter-proximal paused region coincides with accumulation of Pol II upstream of the +1 nucleosome. Together, these observations are consistent with CBP's role in histone acetylation to facilitate paused Pol II release into transcription elongation.

Furthermore, the high-resolution heatmap (Figure C.3B) revealed that the PRO-seq density after CBP inhibition decreases in the first 60 bp for

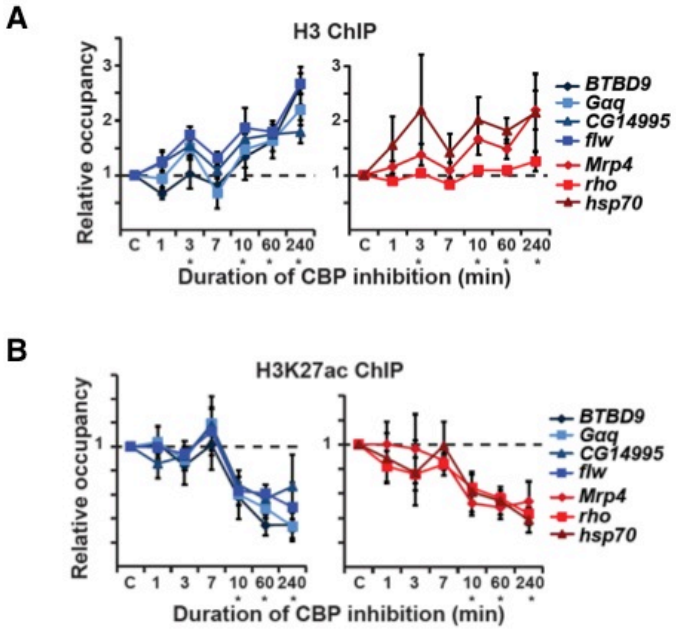


Figure C.4. CBP facilitates productive elongation of Pol II by acetylating nucleosomes

(A) ChIP-qPCR of H3K27ac after CBP inhibitor treatment for the indicated time points (n=2-4). Occupancy at paused and highly-paused promoters are plotted relative to control-treated cells. Error bars represent S.E.M., two-tailed paired t-test, *p<0.05.

(B) ChIP-qPCR of H3 after CBP inhibitor treatment for the indicated time points (n=2-4). Occupancy at paused and highly-paused promoters are plotted relative to control-treated cells. Error bars represent S.E.M., two-tailed paired t-test, *p<0.05.

The data and figure was generated by Ann Boija.

almost all genes, suggesting that CBP is also important for Pol II recruitment at the majority of expressed genes. Taken together, our results reveal that CBP is required for recruitment of Pol II and to allow Pol II to overcome transcriptional barrier caused by +1 nucleosome that resides downstream of the pause region.

Discussion

Elucidating transcriptional regulatory mechanisms is of fundamental importance for understanding cell behavior in both normal and pathological conditions. In metazoan genomes, pausing of Pol II at a position close to the transcription start site is prevalent, occurring on a majority of active genes in a given cell type (Core et al., 2012). How Pol II pausing is regulated is not fully understood. Here we show that only 10 min of CBP inhibition results in dribbling of Pol II from the canonical pause site to positions further downstream. We propose that this effect can be explained by the combination of two different CBP activities, recruitment of Pol II to the promoter and facilitation of transcription elongation beyond the first nucleosome. In this scenario, CBP inhibition causes less Pol II recruitment, resulting in diminished Pol II amounts at the canonical pause site. The Pol II that has initiated transcription is released normally from the canonical pause, but accumulates upstream of the +1 nucleosome. Due to different rate-limiting steps at different

types of promoters, the final outcome of CBP inhibition is either a decrease or an accumulation of Pol II in the promoter-proximal region.

At weakly-paused promoters, most Pol IIs that are recruited are efficiently released into productive elongation. Upon CBP inhibition, paused Pol II dribbles to more downstream positions, accumulates upstream of +1 nucleosome, and fails to effectively enter productive elongation, causing an increase in total Pol II occupancy around the pause region. Thus, CBP is rate-limiting for transcription of these genes at a post-recruitment step, likely by acetylation of the +1 nucleosome. This is consistent with single-cell analyses of transcription induction of a tandem gene array in mouse cells, where H3K27ac and p300/CBP were shown to be required for accumulation of elongating S2P Pol II (Stasevich et al., 2014). Together, these results demonstrate a critical function for CBP in efficient release of Pol II from the promoter-proximal pause into productive elongation. Furthermore, our data suggest that CBP is involved in Pol II recruitment at weakly-paused genes, and therefore stimulates transcription by both Pol II recruitment and release at virtually all promoters. This indicates that CBP activity has its effect at most expressed genes, including those that fall below our cut-off for high confidence CBP ChIP-seq peaks.

Another explanation for the altered positioning of Pol II at the pause site upstream of the first nucleosome could be the CBP-mediated acetylation of some other protein involved in pausing. In mammalian cells, CBP can

acetylate many proteins including the non-consensus lysines in the Pol II CTD (Schroder et al., 2013). Inefficient positioning of paused Pol II upon CBP inhibition could result in insufficient modification of the Pol II CTD for efficient elongation. Yet another possibility is the targeting of pausing factor (e.g. NELF or DSIF) or core promoter factors by CBP.

C646 is the most potent inhibitor of lysine acetyltransferases described, and selectively inhibits p300/CBP over other acetyltransferases (Bowers et al., 2010). Although it is a competitive inhibitor that does not covalently modify p300/CBP (Bowers et al., 2010), covalent off-target reactivity with abundant cellular proteins has also been described (Shrimp et al., 2016). The short exposure (10 min) of the treated cells in our experiments, the use of the similarly reactive compound C37 as a control, and the significant correlation between CBP occupancy and the effects of C646, argue that on-target CBP inhibition accounts for the transcriptional changes observed. Although C646 inhibits CBP's catalytic activity (Bowers et al., 2010), we discovered that occupancy of CBP itself is rapidly affected by the CBP inhibitor, consistent with *in vitro* data showing that a p300/CBP inhibitor causes a conformational change that dissociates p300/CBP from a chromatin template (Black et al., 2006). CBP is also a large protein with many interacting domains and binding partners (reviewed in Bedford et al., 2010). For this reason, we cannot distinguish between CBP-mediated acetylation of a substrate that is required for transcription and a non-enzymatic CBP function.

Our results demonstrate that CBP controls positioning of promoter-proximal paused Pol II at a large fraction of *Drosophila* genes by facilitating transcription through the first nucleosome and by stimulating Pol II recruitment. This global activator function for CBP is consistent with tethering experiments where a Gal4-CBP fusion protein strongly activated transcription from 24 different enhancer contexts (Stampfel et al., 2015). This suggests that CBP has critical roles not only in the previously described control of enhancer activity by H3K27 acetylation (reviewed in Holmqvist and Mannervik, 2013), but also in regulation of Pol II activity at the promoter.

REFERENCES

Adelman, K., and Lis, J.T. (2012). Promoter-proximal pausing of RNA polymerase II: emerging roles in metazoans. Nat Rev Genet *13*, 720-731.

Adkins, N.L., Hagerman, T.A., and Georgel, P. (2006). GAGA protein: a multi-faceted transcription factor. Biochem Cell Biol *84*, 559-567.

Arnold, C.D., Gerlach, D., Stelzer, C., Boryn, L.M., Rath, M., and Stark, A. (2013). Genome-wide quantitative enhancer activity maps identified by STARR-seq. Science *339*, 1074-1077.

Bedford, D.C., Kasper, L.H., Fukuyama, T., and Brindle, P.K. (2010). Target gene context influences the transcriptional requirement for the KAT3 family of CBP and p300 histone acetyltransferases. Epigenetics *5*, 9-15.

Black, J.C., Choi, J.E., Lombardo, S.R., and Carey, M. (2006). A mechanism for coordinating chromatin modification and preinitiation complex assembly. Mol Cell *23*, 809-818.

Bowers, E.M., Yan, G., Mukherjee, C., Orry, A., Wang, L., Holbert, M.A., Crump, N.T., Hazzalin, C.A., Liszczak, G., Yuan, H.*, et al.* (2010). Virtual ligand screening of the p300/CBP histone acetyltransferase: identification of a selective small molecule inhibitor. Chem Biol *17*, 471-482.

Chen, K., Johnston, J., Shao, W., Meier, S., Staber, C., and Zeitlinger, J. (2013). A global change in RNA polymerase II pausing during the Drosophila midblastula transition. Elife *2*, e00861.

Cherbas, L., Willingham, A., Zhang, D., Yang, L., Zou, Y., Eads, B.D., Carlson, J.W., Landolin, J.M., Kapranov, P., Dumais, J.*, et al.* (2011). The transcriptional diversity of 25 Drosophila cell lines. Genome Res *21*, 301-314.

Choi, C.H., Hiromura, M., and Usheva, A. (2003). Transcription factor IIB acetylates itself to regulate transcription. Nature *424*, 965-969.

Core, L.J., and Lis, J.T. (2008). Transcription regulation through promoter-proximal pausing of RNA polymerase II. Science *319*, 1791-1792.

Core, L.J., Waterfall, J.J., Gilchrist, D.A., Fargo, D.C., Kwak, H., Adelman, K., and Lis, J.T. (2012). Defining the status of RNA polymerase at promoters. Cell Rep *2*, 1025-1035.

Crump, N.T., Hazzalin, C.A., Bowers, E.M., Alani, R.M., Cole, P.A., and Mahadevan, L.C. (2011). Dynamic acetylation of all lysine-4 trimethylated histone H3 is evolutionarily conserved and mediated by p300/CBP. Proc Natl Acad Sci U S A *108*, 7814-7819.

Dancy, B.M., Crump, N.T., Peterson, D.J., Mukherjee, C., Bowers, E.M., Ahn, Y.H., Yoshida, M., Zhang, J., Mahadevan, L.C., Meyers, D.J.*, et al.* (2012). Live-cell studies of p300/CBP histone acetyltransferase activity and inhibition. Chembiochem *13*, 2113-2121.

Dennis, G., Jr., Sherman, B.T., Hosack, D.A., Yang, J., Gao, W., Lane, H.C., and Lempicki, R.A. (2003). DAVID: Database for Annotation, Visualization, and Integrated Discovery. Genome Biol *4*, P3.

Duttke, S.H., Lacadie, S.A., Ibrahim, M.M., Glass, C.K., Corcoran, D.L., Benner, C., Heinz, S., Kadonaga, J.T., and Ohler, U. (2015). Human promoters are intrinsically directional. Mol Cell *57*, 674-684.

Enderle, D., Beisel, C., Stadler, M.B., Gerstung, M., Athri, P., and Paro, R. (2011). Polycomb preferentially targets stalled promoters of coding and noncoding transcripts. Genome Res *21*, 216-226.

Feller, C., Forne, I., Imhof, A., and Becker, P.B. (2015). Global and specific responses of the histone acetylome to systematic perturbation. Mol Cell *57*, 559-571.

Fuda, N.J., Guertin, M.J., Sharma, S., Danko, C.G., Martins, A.L., Siepel, A., and Lis, J.T. (2015). GAGA factor maintains nucleosome-free regions and has a role in RNA polymerase II recruitment to promoters. PLoS Genet *11*, e1005108.

Gilchrist, D.A., Dos Santos, G., Fargo, D.C., Xie, B., Gao, Y., Li, L., and Adelman, K. (2010). Pausing of RNA polymerase II disrupts DNA-specified nucleosome organization to enable precise gene regulation. Cell *143*, 540-551.

Hendrix, D.A., Hong, J.W., Zeitlinger, J., Rokhsar, D.S., and Levine, M.S. (2008). Promoter elements associated with RNA Pol II stalling in the Drosophila embryo. Proc Natl Acad Sci U S A *105*, 7762-7767.

Holmqvist, P.H., Boija, A., Philip, P., Crona, F., Stenberg, P., and Mannervik, M. (2012). Preferential genome targeting of the CBP co-activator by Rel and Smad proteins in early Drosophila melanogaster embryos. PLoS Genet *8*, e1002769.

Holmqvist, P.H., and Mannervik, M. (2013). Genomic occupancy of the transcriptional co-activators p300 and CBP. Transcription *4*, 18-23.

Jin, Q., Yu, L.R., Wang, L., Zhang, Z., Kasper, L.H., Lee, J.E., Wang, C., Brindle, P.K., Dent, S.Y., and Ge, K. (2011). Distinct roles of GCN5/PCAF-mediated H3K9ac and CBP/p300-mediated H3K18/27ac in nuclear receptor transactivation. Embo J *30*, 249-262.

Kubik, S., Bruzzone, M.J., Jacquet, P., Falcone, J.L., Rougemont, J., and Shore, D. (2015). Nucleosome Stability Distinguishes Two Different Promoter Types at All Protein-Coding Genes in Yeast. Mol Cell *60*, 422-434.

Kwak, H., Fuda, N.J., Core, L.J., and Lis, J.T. (2013). Precise maps of RNA polymerase reveal how promoters direct initiation and pausing. Science *339*, 950-953.

Kwok, R.P., Lundblad, J.R., Chrivia, J.C., Richards, J.P., Bachinger, H.P., Brennan, R.G., Roberts, S.G., Green, M.R., and Goodman, R.H. (1994). Nuclear protein CBP is a coactivator for the transcription factor CREB. Nature *370*, 223-226.

Lee, C., Li, X., Hechmer, A., Eisen, M., Biggin, M.D., Venters, B.J., Jiang, C., Li, J., Pugh, B.F., and Gilmour, D.S. (2008). NELF and GAGA factor are linked to promoter-proximal pausing at many genes in Drosophila. Mol Cell Biol *28*, 3290-3300.

Lilja, T., Aihara, H., Stabell, M., Nibu, Y., and Mannervik, M. (2007). The acetyltransferase activity of Drosophila CBP is dispensable for regulation of the Dpp pathway in the early embryo. Dev Biol *305*, 650-658.

Negre, N., Brown, C.D., Ma, L., Bristow, C.A., Miller, S.W., Wagner, U., Kheradpour, P., Eaton, M.L., Loriaux, P., Sealfon, R.*, et al.* (2011). A cis-regulatory map of the Drosophila genome. Nature *471*, 527-531.

Philip, P., Boija, A., Vaid, R., Churcher, A.M., Meyers, D.J., Cole, P.A., Mannervik, M., and Stenberg, P. (2015). CBP binding outside of promoters and enhancers in. Epigenetics Chromatin *8*, 48.

Roy, S., Ernst, J., Kharchenko, P.V., Kheradpour, P., Negre, N., Eaton, M.L., Landolin, J.M., Bristow, C.A., Ma, L., Lin, M.F.*, et al.* (2010). Identification of functional elements and regulatory circuits by Drosophila modENCODE. Science *330*, 1787-1797.

Schneider, I. (1972). Cell lines derived from late embryonic stages of Drosophila melanogaster. J Embryol Exp Morphol *27*, 353-365.

Schroder, S., Herker, E., Itzen, F., He, D., Thomas, S., Gilchrist, D.A., Kaehlcke, K., Cho, S., Pollard, K.S., Capra, J.A.*, et al.* (2013). Acetylation of RNA polymerase II regulates growth-factor-induced gene transcription in mammalian cells. Mol Cell *52*, 314-324.

Shrimp, J.H., Sorum, A.W., Garlick, J.M., Guasch, L., Nicklaus, M.C., and Meier, J.L. (2016). Characterizing the Covalent Targets of a Small Molecule Inhibitor of the Lysine Acetyltransferase P300. ACS Med Chem Lett *7*, 151-155.

St Pierre, S.E., Ponting, L., Stefancsik, R., and McQuilton, P. (2014). FlyBase 102--advanced approaches to interrogating FlyBase. Nucleic Acids Res *42*, D780-788.

Stampfel, G., Kazmar, T., Frank, O., Wienerroither, S., Reiter, F., and Stark, A. (2015). Transcriptional regulators form diverse groups with context-dependent regulatory functions. Nature *528*, 147-151.

Stasevich, T.J., Hayashi-Takanaka, Y., Sato, Y., Maehara, K., Ohkawa, Y., Sakata-Sogawa, K., Tokunaga, M., Nagase, T., Nozaki, N., McNally, J.G.*, et al.* (2014). Regulation of RNA polymerase II activation by histone acetylation in single living cells. Nature *516*, 272-275.

Tie, F., Banerjee, R., Stratton, C.A., Prasad-Sinha, J., Stepanik, V., Zlobin, A., Diaz, M.O., Scacheri, P.C., and Harte, P.J. (2009). CBP-mediated acetylation of histone H3 lysine 27 antagonizes Drosophila Polycomb silencing. Development *136*, 3131-3141.

Wang, Z., Zang, C., Cui, K., Schones, D.E., Barski, A., Peng, W., and Zhao, K. (2009). Genome-wide mapping of HATs and HDACs reveals distinct functions in active and inactive genes. Cell *138*, 1019-1031.

Zhou, Q., Li, T., and Price, D.H. (2012). RNA polymerase II elongation control. Annu Rev Biochem *81*, 119-143.